

ATTN: Ed. Hart.

CLFE

Access DB# 103454

RECEIVED

# SEARCH REQUEST FORM

SEP -9 2003

Scientific and Technical Information Center

STIC/BIOTECH DIVISION  
(STIC)

Requester's Full Name: Celine Qian Examiner #: 78710 Date: 9/8/03  
Art Unit: 1636 Phone Number 306-0283 Serial Number: 091877935  
Mail Box and Bldg/Room Location: 11E01/11C10 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Regulatory seq of mouse villin gene  
Inventors (please provide full names): D. Pinto et al.

Earliest Priority Filing Date: 6/8/2001

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search for SEQ ID NO: 1 & oligo search.

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

## STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone #: \_\_\_\_\_  
Searcher Location: \_\_\_\_\_  
Date Searcher Picked Up: 9/12/03  
Date Completed: 9/15/03  
Searcher Prep & Review Time: \_\_\_\_\_  
Clerical Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

## Type of Search

NA Sequence (#) 2  
AA Sequence (#) \_\_\_\_\_  
Structure (#) \_\_\_\_\_  
Bibliographic \_\_\_\_\_  
Litigation \_\_\_\_\_  
Fulltext \_\_\_\_\_  
Patent Family \_\_\_\_\_  
Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_  
Dialog \_\_\_\_\_  
Questel/Orbit \_\_\_\_\_  
Dr.Link \_\_\_\_\_  
Lexis/Nexis \_\_\_\_\_  
Sequence Systems 04/03  
WWW/Internet \_\_\_\_\_  
Other (specify) \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**.

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**.

*The Pending database search results should not be left in the case because they contain data that is confidential.*

**THIS PAGE BLANK (USPTO)**





# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 103454**

**TO: Celine Qian**  
**Location: CM-1/11C01/11E12**  
**Art Unit: 1636**  
**Monday, September 15, 2003**  
  
**Case Serial Number: 09/877935**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**CM1-6B02**  
**Phone: 305-9203**  
  
**edward.hart@uspto.gov**

### **Search Notes**

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

**THIS PAGE BLANK (USPTO)**



# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact:*

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 19:07:58 : Search time 16404 Seconds  
(without alignments)  
13327.162 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctgtgcacacagagaca.....ctctagctctgcacacatg 8995

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	498	5.5	507	28	AZ354198 1M0093E23
2	445.6	5.0	495	28	BH086621 RPT-24-2
3	322.6	3.6	336	28	A2657211 1M0533615
4	291.8	3.2	307	28	AZ354559 1M0093E23

C 5	280	3.1	291	28	AZ821333
C 6	257.2	2.9	639	28	AZ821403
C 7	176.4	2.0	641	10	BB612611
C 8	175.8	2.0	589	14	CA884544
C 9	175.8	2.0	4682	11	AK052834
C 10	174	1.9	423	14	CB178219
C 11	173.4	1.9	399	12	B1465446
C 12	172.4	1.9	383	12	BM875096
C 13	170.4	1.9	747	9	AV307409
C 14	166.8	1.9	336	13	BY324593
C 15	166.8	1.9	346	13	BY327420
C 16	166.4	1.8	449	9	A1836762
C 17	166.4	1.8	3090	11	AK037096
C 18	165.4	1.8	703	28	BH108871
C 19	165.2	1.8	510	28	A2406837
C 20	165.2	1.8	1491	11	BC024585
C 21	164.6	1.8	682	28	A2313191
C 22	162.6	1.8	502	14	CA947391
C 23	161	1.8	623	28	A2766497
C 24	161	1.8	626	28	A2369208
C 25	161	1.8	634	28	A2564931
C 26	160.6	1.8	630	28	A2579634
C 27	160	1.8	558	13	BQ478801
C 28	159.2	1.8	479	9	AA238960
C 29	159.2	1.8	666	10	BG670252
C 30	159	1.8	538	28	A2699259
C 31	158.4	1.8	630	10	BB655081
C 32	158.4	1.8	782	28	BH069206
C 33	157.8	1.8	637	10	BB620553
C 34	157.8	1.8	897	11	AK077869
C 35	157.6	1.8	409	28	A2075194
C 36	157.6	1.8	504	28	A2980279
C 37	157.4	1.7	409	14	CA879083
C 38	157.4	1.7	518	28	A2034666
C 39	157.2	1.7	590	28	A2645050
C 40	157.2	1.7	3874	11	AK083490
C 41	156.4	1.7	607	14	BY721781
C 42	156.2	1.7	424	10	BB673144
C 43	156.2	1.7	745	12	BM933152
C 44	156.2	1.7	2114	11	AK008714
C 45	156	1.7	627	28	A2809859

#### ALIGNMENTS

RESULT 1  
LOCUS AZ354198/c 507 bp DNA linear GSS 02-OCT-2000  
DEFINITION 1M0093E23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0093E23 F, genomic survey sequence.  
ACCESSION AZ354198  
VERSION AZ354198.1 GI:10465402  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 507)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0093 row: E column: 23  
Seq primer: CGTTGTAAACGACGCGCCACT  
Class: plasmid ends  
High quality sequence stop: 507.

## FEATURES

## SOURCE

Location/Qualifiers

1. 507  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCIM0093E23"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_1lb="Mouse 10kb plasmid UUCIM library"  
/note="Vector: PMD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 125 a 139 c 105 g 138 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 498; DB 28; Length 507;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8431 CATCCAGAGTGGCTAGCCAGGAGACTCAGGATATGCTGCTCTCCCTCCACAGTCT 8490

507 CATCCAGAGTGGCTAGCCAGGAGACTCAGGATATGCTGCTCTCCCTCCACAGTCT 448

8491 AGAATTACAGGACATACATCTGCTGGAATTTTAACTGAACTCTGAGATAGCA 8550

447 AGAATTACAGGACATACATCTGCTGGAATTTTAACTGAACTCTGAGATAGCA 388

8551 GGCACCTCTCAATGAGAGGCTCTTTTGTGTTGGTTGGTTTCTCTCATTAAGTCA 8610

387 GGCACCTCTCAATGAGAGGCTCTTTTGTGTTGGTTGGTTTCTCTCATTAAGTCA 328

8611 GGCAGCTGTAATAGTCTAGCGGCTAGATTAACATCTGCTCAAAAAGCCATAGAG 8670

327 GGCAGCTGTAATAGTCTAGCGGCTAGATTAACATCTGCTCAAAAAGCCATAGAG 268

8671 GTAGGAGGTGAGAGGCTTAAGAGAGCCGCTGATATGACACAGAGATAGCC 8730

267 GTAGGAGGTGAGAGGCTTAAGAGAGCCGCTGATATGACACAGAGATAGCC 208

8731 TGCACATATATAGCAAGACCTTGTTCAAAAACATGAGGAGGAGGTATGTTTAAAGTCT 8790

207 TGCACATATATAGCAAGACCTTGTTCAAAAACATGAGGAGGAGGTATGTTTAAAGTCT 148

8791 GGGCTGTGTAAAGGACCTAAGGAGGACATGTAGACATTTGACTAAGAAAGATCATCA 8850

147 GGGCTGTGTAAAGGACCTAAGGAGGACATGTAGACATTTGACTAAGAAAGATCATCA 88

8851 TCAAAGCCGGGTGGGAGGTAGAGGTGAGCTACAGTGGTCAAGAGCCCATATAGGAAG 8910

87 TCAAAGCCGGGTGGGAGGTAGAGGTGAGCTACAGTGGTCAAGAGCCCATATAGGAAG 28

QY 8911 CAGTTCCCTCTCTCTC 8928  
Db 27 CAGTTCCCTCTCTCTC 10

## RESULT 2

## LOCUS

BH086621 495 bp DNA linear GSS 18-Jul-2001

DEFINITION RPI-24-263015.TJ RPI-24 Mus musculus genomic clone RPI-24-263015

, genomic survey sequence.

ACCESION BH086621 GI:14906388

VERSION BH086621.1 GI:14906388

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus.

1 (bases 1 to 495)

Zhao, S., Nierman, W., Malek, J., Shatsman, S., Aklnet, B., Levins, M.,

Teagaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,

Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPI-24

Unpublished

Other\_GSSs: RPI-24-263015.TVB

COMMENT Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@tigr.org). Clones may be purchased from BACPAC

Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 263 row: 0 column: 15

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 495

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPI-24-263015"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pTRABAC1; site\_1: BamHI; site\_2: BamHI;

RPI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTRABAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 107 a 148 c 153 g 87 t

ORIGIN

Query Match

Best Local Similarity 97.6%; Score 445.6; DB 28; Length 495;

Matches 484; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

2740 GGAAGCAGGCGCATGTTTCCA-GAGACCTACAGCAGAGGCGAGCAAGAGATCCCGAG 2798

1 GGAAGCAGGCGCATGTTTCCAATTTCATACAGCAGAGGCGAGCAAGAGATCCCGAG 60

2799 GTCCAGGCGCAGGAGGTGAGGCGCTTGTTCGAGAGGAGGCGGCGAGCAAGAGGT 2858

61 GTCCAGGCGCAGGAGGTGAGGCGCTTGTTCGAGAGGAGGCGGCGAGCAAGAGGT 120

2859 TCAAAGCAGCAGGTTTATGAGCTCATTAAGAGGAGGTCGAGGCTCAGTCAAGAAAGGA 2918

121 TCAAAGCAGCAGGTTTATGAGCTCATTAAGAGGAGGTCGAGGCTCAGTCA-TAAGTC 179



Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0093 row: E column: 23  
Seq primer: CACACAGAAACAGCATATGACC  
Class: plasmid ends  
High quality sequence stop: 307.

# FEATURES

## source

1. 307  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U09C1M0093E23"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid U09C1M library"  
/note="Vector: PWD42nv, Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321149b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 63 a 67 c 83 g 94 t  
ORIGIN

Query Match 3.2%; Score 291.8; DB 28; Length 307;  
Best Local Similarity 99.0%; Pred. No. 1.1e-21;  
Matches 304; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1084 GCCAAGCGTGGTCTCTGCAAGAC-AGCCAATGCGCTTAACCATGGACCGACCTCT 1142  
DB 1 GCCAAGCGTGGTCTCTGCAAGACATCTAGCCAGCTTACCCCTGGGACGCTCTCT 60  
QY 1143 AGCCTAAGGTATCTTTAGTTTAAAAATATATATCTCAGCCGGGTGCTGGCAC 1202  
DB 61 AGCCTAAGGTATCTTTAGTTTAAAAATATATATCTCAGCCGGGTGCTGGCAC 120  
QY 1203 ACGCCCTTAATCCACACACTTGAGAGCTGAGTGAATTAACACACAGCCAGCT 1262  
DB 121 AGCCCTTAATCCACACACTTGAGAGCTGAGTGAATTAACACACAGCCAGCT 180  
QY 1263 GGGGTGACAGCTGGCCCTGTTTGTCTTTCTTATATGCACTGGTCTTACC 1322  
DB 181 GGGGTGACAGCTGGCCCTGTTTGTCTTTCTTATATGCACTGGTCTTACC 240  
QY 1323 TGGGTATGTCCTGCAAGGCTGATCCCTGGAGCTGGAGTTAAAGACAGTTGTG 1382  
DB 241 TGGGTATGTCCTGCAAGGCTGATCCCTGGAGCTGGAGTTAAAGACAGTTGTG 300  
QY 1383 ATCAGCG 1389  
DB 301 ATCAGCG 307

RESULT 5  
A2821333/c 291 bp DNA linear GSS 20-FEB-2001  
LOCUS A2821333  
DEFINITION 2M0094K06F Mouse 10kb plasmid U09C1M library Mus musculus genomic  
clone U09C2M0094K06 F, genomic survey sequence.

ACCESSION A2821333  
VERSION A2821333.1 GI:12991241  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weis,R.  
1 (bases 1 to 291)

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0094 row: K column: 06  
Seq primer: CGTTGTAACAGCAGCGCCACT  
Class: plasmid ends  
High quality sequence stop: 291.

## FEATURES

## source

1. 291  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U09C2M0094K06"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid U09C1M library"  
/note="Vector: PWD42nv, Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321149b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 40 a 89 c 85 t  
ORIGIN

Query Match 3.1%; Score 280; DB 28; Length 291;  
Best Local Similarity 99.7%; Pred. No. 2.1e-20;  
Matches 291; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2583 ACACACACACACACACACACATTCAGTCTCAGAGCTCTTGGAGGTCACG 2642  
DB 291 ACACACACACACACACACACATTCAGTCTCAGAGCTCTTGGAGGTCACG 232  
QY 2643 AAGAGCTGCGCTCAACACAGATCTTCATCTTCCCTCTAAAGAGACACAGATTCCAA 2702  
DB 231 AAGAGCTGCGCTCAACACAGATCTTCATCTTCCCTCTAAAGAGACACAGATTCCAA 172  
QY 2703 GGTGGCAGAAAGATCTACAGGGGCGACAGGAGGAGGAGGAGCAGGCCATGTTCACG 2762







[illegible]

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
DB	2077	GCCCTTAATCCACGACCTTGGGAGGACAGGACAGGTGATCTCTGATTCGAGGCCAGCC										
QY	4668	TGCTCTAATAGAGTGTGATTCGAGACAGGACGCGCTACACAGAGAAACCCCTGTTTGAAAA	4722									
DB	1967	TGCTCTAATAGAGTGTGATTCGAGACAGGACGCGCTACACAGAGAAACCCCTGTTTGAAAA	1908									
QY	4728	ACCAAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	4774									
DB	1907	ACCAAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	1861									
RESULT 10	CB178219/c											
LOCUS	CB178219	423 bp	mRNA	linear	EST 31-JAN-2003							
DEFINITION	in5b04.y1 Kaestner ngm3 wt	Mus musculus	cDNA 5'	mRNA sequence.								
ACCESSION	CB178219											
VERSION	CB178219.1	GI:28186609										
KEYWORDS	EST.											
SOURCE	Mus musculus											
ORGANISM	Mus musculus (house mouse)											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.											
AUTHORS	1 (bases 1 to 423)											
	Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K., Lemisha, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, J., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.											
	Endocrine Pancreas Consortium											
	Unpublished											
	Other_ESTs: in5b04.x1											
	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue											
	Endocrine Pancreas Consortium											
	Harvard University, Howard Hughes Medical Institute											
	Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138											
	Tel: 617-495-1812											
	Fax: 617-495-8557											
	Email: dmelton@biohp.harvard.edu											
	Pancreas was obtained from Gerard Gradwohl (PNAS 97 p1607-1611, 2000) library was constructed by Catherine Lee DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: Dr. Marie Scearce (mscearce@mail.med.upenn.edu)											
	Seq primer: -408P from GIBCO											
	High quality sequence stop: 398.											
FEATURES	Location/Qualifiers											
Source	1..423											
	/organism="Mus musculus"											
	/mol_type="mRNA"											
	/strain="129/Sv x CD1"											
	/db_xref="taxon:10090"											
	/lab_stage="P.c. 14.5"											
	/dev_host="E. coli-DH12S (GIBCO)"											
	/clone_lib="Kaestner ngm3 wt"											
	/note="Organ: pancreas; Vector: pSPORT1 (GIBCO); Site:1: Not 1; Site:2: Sal I; The library was prepared by Catherine S. Lee and has not been published. The pancreas was obtained from Gerard Gradwohl (PNAS 97 p1607-1611, 2000). The cDNA's were prepared with an oligo containing a NotI site, and SalI linkers were added to the ends. The NotI-SalI sites in the vectors. This is one of two libraries, ngm3 wt and ngm3 -/- . The wt library is in pSPORT1, T7 promoter is 5'."											
BASE COUNT	90 a	107 c	89 g	137 t								
ORIGIN					</							

OY		4489	TAACTGAGCACTTGGGGGGCGGTGAAGGAG-CATTTCCTATTAATTTGAGCCAGCCTG	454.7
Db		323	TATTCCACGACGCTGGGAAGCTGACACAGACTTACCCTAAGAAGTTCAAGCTTAGCCTG	264
OY		4548	AGCTATAGAGCGAGACTTGTCTTTAAAGAAAAAATGMAAGCCACGACAGTGGTGGACAC	4607
Db		263	GCGTACACAGCAGGCCCTGTCTCAGAGAAATAATATGCTGTGGCGCGGTGGCGCAC	204
OY		4608	GCCCTTAATCCAGCACTTTGGGAGGCAGACAGCAGATTTCTGTAGTTCAAGGCCAGCC	4667
Db		203	GCCCTTAATCCACACACTTTGGGAGGCAGACAGCAGAGTGATCTGTAGTTCCAGGCCAGCC	144
OY		4668	TGGCTCTATAGAGTGGTTCACGAGCAGCGGCTACACAGAAAACCCCTTTTGA AAA	4727
Db		143	TGGCTCTACAGAGTGTGATTCACGAGCAGCCAGCACTACACAGAAAACCCCTTTTGA AAA	84
OY		4728	ACCAGAAAAACAACAAACAAACAAACAAACAA 4761	
Db		83	CCCCAAAAAAAAAAAAAAAAAGAAAGAAAAAAA 50	
RESULT 11				
B1465446/c				
LOCUS		B1465446	399 bp	mRNA linear EST 22-AUG-2000
DEFINITION		iel7c03.y1	Kaestner ngn3 wt	Mus musculus cDNA 5', mRNA sequence.
ACCESSION		B1465446		
VERSION		B1465446.1	GI:15277953	
KEYWORDS		EST.		
SOURCE		Mus musculus (house mouse)		
ORGANISM		Mus musculus		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
JOURNAL		1 (bases 1 to 399)		
COMMENT		Melton,D., Brown,J., Kenty,G., Permult,A., Lee,C., Kaestner,K., Lemisha,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blisstein,A., Schmitt,A., Theising,B., Rither,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McMan,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.		
		Endocrine Pancreas Consortium		
		Unpublished		
		Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue		
		Harvard University, Howard Hughes Medical Institute		
		Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,		
		MA 02138		
		Tel: 617-495-1812		
		Fax: 617-495-8557		
		Email: dmelton@lhop.harvard.edu		
		Pancreas was obtained from Gerard Gradwohl (PMAS 97 P1607-1611,		
		2000) Library was constructed by Catherine Lee DNA sequencing by:		
		Washington University Genome Sequencing Center for information on		
		obtaining a clone please contact: Dr. Marie Scarce		
		(mscarce@mail.med.upenn.edu)		
		High quality sequence stop: 268.		
FEATURES				
SOURCE				
		location/Qualifiers		
		1..399		
		/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="129/Sv x CD1"		
		/db_xref="taxon:10090"		
		/dev_stage="P.C. 14.5"		
		/lab_host="E. coli DH12S (GIBCO)"		
		/clone_id="Kaestner ngn3 wt"		
		/note="Organ: pancreas; Vector: pSPOR1 (GIBCO); Site_1:		
		Not I; Site_2: Sal I; The library was prepared by		
		Catherine S. Lee and has not been published. The pancreas		
		was obtained from Gerard Gradwohl (PMAS 97 P1607-1611,		
		2000). The cDNA's were prepared with an oligo containing a		
		NotI site, and SalI linkers were added to the ends. The		
		inserts were cut with NotI before being cloned into the		
		Not-SalI sites in the vectors. This is one of two		





prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Vassilis Aidinis ( Biomedical Sciences Research Center 'Al. Fleming' Institute of Immunology 14-16 Al. Fleming street 16672 Vari, Greece ) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
source Location/Qualifiers

1..336  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="L030024J05"  
/cell\_type="synovial fibroblasts"  
/clone\_lib="RIKEN full-length enriched, synovial fibroblasts"

BASE COUNT 70 a 86 c 76 g 104 t

ORIGIN

Query Match

1.9% Score 166.8; DB 13; Length 336;

Best Local Similarity 79.8%; Pred. No. 1.8e-08;

Matches 209; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 4489 TAATCTGACACTTGGAGGCTGAAGCAGAG-GATTCTATATTTGGAGCCAGCCTG 4547

DB 287 TATTCACAGAGCTGGAAAGCTGACAAAGATTAACATTAAGACTTCAAGGCTGACCTG 228

QY 4548 AGCTTACAGCAGACTTGTCTTAAAGAAAAAATGAAGCCAGCAGTGTGGCAGC 4607

DB 227 GCTACACAGCCGCGCTGTCTCAAGAAATTAATGCTGGCGGTGGCGC 168

QY 4608 GCCTTAATCCAGACTTGGAGGAGCAGACGAGATTTCTAGTTCAAGCCAGCC 4667

DB 167 GCCTTAATCCAGACTTGGAGGAGCAGACGAGTGTCTCTAGTTCCAGCCAGCC 108

QY 4668 TGTCTATAGAGTGAAGTCCAGACAGCCAGGCTTACACAGAAACCCGTTTGAANA 4727

DB 107 TGTCTACAGAGTGAAGTCCAGACAGCCAGGCTTACACAGAAACCCGTTTGAANA 48

QY 4728 ACCAGAAAAACAACAAACA 4749

DB 47 ACCAGAAAAACAACAAACA 26

RESULT 15  
BY327420/c 346 bp mRNA linear EST 11-DEC-2002

DEFINITION BY327420 RIKEN full-length enriched, synovial fibroblasts Mus

musculus cDNA clone L030041P18 5', mRNA sequence.

ACCESSION BY327420

VERSION BY327420.1 GI:26517987

KEYWORDS EST

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 346)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamataka, I., Kiyosawa, H.,

Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,

Gojobori, T., Baldarelli, R., Hall, D.P., Bult, C., Hume, D.A.,

TITLE  
JOURNAL MEDLINE PUBMED  
COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>,  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,  
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Waki, K., Watanabe, A., Yamashita, M., and Hayashizaki, Y. Direct  
Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapped-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN Integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Vassilis Aidinis ( Biomedical Sciences  
Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.  
Fleming street 16672 Vari, Greece ) whose assistance we gratefully  
acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

FEATURES  
source Location/Qualifiers

1..346  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="L030041P18"  
/cell\_type="synovial fibroblasts"  
/clone\_lib="RIKEN full-length enriched, synovial  
fibroblasts"





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 18:15:33 ; Search time 438 Seconds  
(without alignments)  
9064.478 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995  
Sequence: 1 gatctggtgcacaaagaca.....ctctagctcgtccacatg 8995

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/1na/PTCUTS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/1na/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148.8	1.7	7208	3	US-09-166-186-107
2	148.8	1.7	7208	3	US-09-313-932-107
3	142.6	1.6	14707	4	US-09-312-762A-3
4	140.4	1.6	3481	3	US-08-965-729A-1
5	138.2	1.5	17056	3	US-09-245-041-3
6	138	1.5	5973	3	US-09-245-041-4
7	136.8	1.5	2509	4	US-09-319-284-1
8	135.4	1.5	90050	3	US-09-245-041-5
9	135.2	1.5	4698	1	US-07-807-043B-5
10	135.2	1.5	4698	1	US-08-299-849B-5
11	135.2	1.5	4698	2	US-08-142-368A-5
12	135.2	1.5	4698	3	US-08-967-727-5
13	135.2	1.5	4698	3	US-08-037-230D-5
14	135.2	1.5	4698	4	US-09-583-850-5
15	135.2	1.5	4698	4	US-09-579-197-5
16	135.2	1.5	4698	4	US-09-404-026-5
17	134	1.5	6727	3	US-08-629-643A-5
18	134	1.5	6727	3	US-09-280-799-1
19	134	1.5	6727	3	US-09-155-884-5
20	133.4	1.5	51259	3	US-08-781-891-209
21	133.4	1.5	51259	4	US-09-618-166-209
22	133.8	1.5	2574	4	US-09-668-313A-10
23	133.8	1.5	17056	3	US-09-245-041-3
24	133.2	1.5	48974	3	US-08-920-422-17
25	131.8	1.5	4550	3	US-09-338-907-182
26	131.8	1.5	4550	3	US-09-218-207-182
27	131.4	1.5	3892	2	US-08-555-723B-3

28	131.4	1.5	3892	3	US-09-123-465-3	Sequence 3, Appl1
29	130.8	1.5	90050	3	US-09-245-041-5	Sequence 5, Appl1
30	130.6	1.5	37950	4	US-09-338-907-183	Sequence 183, App
31	130.6	1.5	37950	4	US-09-218-207-183	Sequence 183, App
32	129.6	1.4	35828	4	US-09-449-218D-17	Sequence 17, Appl
33	129.6	1.4	35828	4	US-09-668-529A-17	Sequence 17, Appl
34	129.6	1.4	35828	4	US-09-618-166-209	Sequence 17, Appl
35	129.2	1.4	14707	4	US-09-312-762A-3	Sequence 3, Appl1
36	129	1.4	7874	4	US-09-780-175-96	Sequence 96, Appl
37	127.8	1.4	48974	3	US-08-920-422-17	Sequence 17, Appl
38	126.4	1.4	3217	3	US-09-232-200-64	Sequence 64, Appl
39	126.4	1.4	3217	4	US-09-232-197-64	Sequence 64, Appl
40	126.4	1.4	3217	4	US-09-232-201-64	Sequence 64, Appl
41	125.6	1.4	5687	2	US-08-380-403A-3	Sequence 3, Appl1
42	125.6	1.4	5687	2	US-08-895-628-3	Sequence 3, Appl1
43	125.6	1.4	5687	4	US-08-895-810D-3	Sequence 3, Appl1
44	123.6	1.4	1355	1	US-07-757-390-4	Sequence 4, Appl1
45	123.6	1.4	1355	1	US-07-757-390-16	Sequence 16, Appl1

ALIGNMENTS

RESULT 1  
US-09-166-186-107  
Sequence 107, Application US/09166186A  
Patent No. 6080580  
GENERAL INFORMATION:  
APPLICANT: Baker, Brenda  
APPLICANT: Bennett, C. Frank  
APPLICANT: Butler, Madeline M.  
APPLICANT: Shanahan, William R.  
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE MODULATION OF TNF- $\alpha$  EXPRESSION  
FILE REFERENCE: ISPH-0322  
CURRENT APPLICATION NUMBER: US/09/166,186A  
CURRENT FILING DATE: 1998-10-05  
NUMBER OF SEQ ID NOS: 250  
SEQ ID NO 107  
LENGTH: 7208  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (4527..4712,5225..5279,5457..5504,5799..6217)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (4371)..(4712)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (4713)..(5224)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5225)..(5279)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (5280)..(5456)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5457)..(5504)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (5505)..(5798)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5799)..(6972)  
PUBLICATION INFORMATION:  
AUTHORS: Semon, D.  
AUTHORS: Kawashima, E.  
AUTHORS: Jongeneel, C.V.  
AUTHORS: Shakhov, A.N.  
AUTHORS: Nedospasov, S.A.  
TITLE: Nucleotide sequence of the murine TNF locus, including the  
TITLE: TNF-alpha (tumor necrosis factor) and TNF-beta (lymphotoxin) genes



```

; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELER:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14707
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-312-762A-3

Query Match      1.6%; Score 142.6; DB 4; Length 14707;
Best Local Similarity 75.1%; Pred. No. 2.4e-25;
Matches 178; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 4576 AAAAAATGAAGCCAGAGGCGGACACAGCCTTAATCCAGCCTTGGAGGCG 4635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12047 AAGAGTTAAAGCGCGGCGGATGATGACACGCTTTATCCAGCTTGGAGGCG 12106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4636 AAGCAGCAGATTCTGAGTTCAAGCCAGCCTGCTATAGAGTTCAGAGCAGC 4695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12107 AGCGCGGCGGATTTGAGTTGAGGCGCAGCTGTATCAAAAGAGTTCCAGAGCAGC 12166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4656 CAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAAACAAACAAACAA 4755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12167 CAGGCTACACAGAGAAACCTGTTTGAAGAAACAAACAAACAAACAA 12226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4756 AACAAGCCAAACCAACCAACCTCTCATCTCTAGCTGTGCT 4812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12227 GTTGAAGACAGTAGAAGGCTCCCTCTGCTGCGCCGAGCTAGGTTTGCTT 12283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
; US-08-965-729A-1/c
; Sequence 1, Application US/08965729A
; Patent No. 6200751
; GENERAL INFORMATION:
; APPLICANT: Jian-Ming Gu and Charles T. Esmon
; TITLE OF INVENTION: ENDOTHELIAL SPECIFIC EXPRESSION
; TITLE OF INVENTION: REGULATED BY EPOR CONTROL ELEMENTS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,729A
; FILING DATE: 07-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,718
; FILING DATE: 08-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 164 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: /note="Nucleotides 2270 through 2840 are a
; OTHER INFORMATION: large endothelial specific element"; murine
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: /note="Nucleotides 2990 through 3061 are a
; OTHER INFORMATION: serum response element"; murine
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: /note="Nucleotides 3007 through 3014 are a
; OTHER INFORMATION: thrombin responsive element"; murine
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: /note="Nucleotides 3130 through 3350 are an
; OTHER INFORMATION: endothelial specific element"; murine
; US-08-965-729A-1

Query Match      1.6%; Score 140.4; DB 3; Length 3481;
Best Local Similarity 74.8%; Pred. No. 4.2e-25;
Matches 202; Conservative 0; Mismatches 66; Indels 2; Gaps 2;

QY 4564 TTTGCTTTAAGAAAAATGAAGCCAGCAGTGTGACACAGCCTTAATCCAGCA 4623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2563 TTCACTTTAAAAAAGCCGCGGCTGTGACACAGCCTTAATCCAGCA 2504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4624 CTGGAGAGCAGACAGCAGATTCTGAGTTCAAGCCAGCCTGCTATAGAGTGAG 4683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2503 CTGGAGAGCAGACAGCAGATTCTGAGTTCAAGCCAGCCTGCTATAGAGTGAG 2444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4684 TTCCAGAGCAGCAGGCTACACAGAAACCTGTTGAAGAAACCAAA 4742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2443 -TCAGAGACAGCAGGCTACACAGAAACCTGTTGAAGAAACCAAA 2385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4743 CAAACCAAAACCAAAACCAACCAACCAACCTCTCATCTCTCTCTA 4802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2384 AAAAAACCAAAACCAACCAACCAACCAACCAACCAACCAACCA 2325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4803 GCTGTCTCTCTAGGTGAGTGTG 4832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2324 AAGCAGGCTCTACTATCATATAGAGTTGG 2295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
; US-09-245-041-3
; Sequence 3, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 17056
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-245-041-3

Query Match      1.5%; Score 138.2; DB 3; Length 17056;
Best Local Similarity 79.2%; Pred. No. 3.3e-24;
Matches 164; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```



```

RESULT 9
US-07-807-043B-5/C
; Sequence 5, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-07-807-043B-5

Query Match      1.5%; Score 135.2; DB 1; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCATGCTGGCAGACGCTTTAATCCGACACTTGGGAGGAGGAGGAGGAGATTCTG 4652
DB 1706 GCACTCTGCTGGCAGGCTTTAATCCGACACTTGGGAGGAGGAGGAGGAGATTCTG 1647
QY 4653 AGTTCAGGCGCAGCGCTGTATAGAGTATCCAGACAGCGCGGCTCAGCAGAGAA 4712
DB 1646 AGTTGAGGCGCAGCGCTGTATAGAGTATCCAGACAGCGCGGATCAAGAGAA 1587
QY 4713 ACCCTGTTTGAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAGAA 4764
DB 1586 ACCCTGCTCGAAAAACCAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAGGCG 1535

RESULT 10
US-08-299-849B-5/C
; Sequence 5, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Faljeur, Thierry;
; APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;

```

```

; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9-JULY-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-299-849B-5

Query Match      1.5%; Score 135.2; DB 1; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCATGCTGGCAGACGCTTTAATCCGACACTTGGGAGGCGGAGGAGGAGATTCTG 4652
DB 1706 GCACTCTGCTGGCAGGCTTTAATCCGACACTTGGGAGGCGGAGGAGGAGATTCTG 1647
QY 4653 AGTTCAGGCGCAGCGCTGTATAGAGTATCCAGACAGCGCGGCTCAGCAGAGAA 4712
DB 1646 AGTTGAGGCGCAGCGCTGTATAGAGTATCCAGACAGCGCGGATCAAGAGAA 1587
QY 4713 ACCCTGTTTGAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAGAA 4764
DB 1586 ACCCTGCTCGAAAAACCAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAGGCG 1535

RESULT 11
US-08-142-368A-5/C
; Sequence 5, Application US/08142368A

```

```

; Patent No. 5925729
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/142,368A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5925729man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-142-368A-5

Query Match 1.5%; Score 135.2; DB 2; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCAGTGTGGCAGACGCGCTTTATCCACACTTGGAGGAGCAGAAACAGGCAAGATTTCG 4652
DB 1706 GCAGTGTGGCAGACGCGCTTTATCCACACTTGGAGGAGCAGAAACAGGCAAGATTTCG 1647
QY 4653 AGTTCAAGCCAGCCGCTGTATAGAGTTCACGAGCAGGCGCTACACAGAGAA 4712
DB 1646 AGTTTGAAGCCAGCCGCTGTACAGAGTGTCCAGGACAGGCGATACAAAGAAA 1587
QY 4713 ACCCTGTTTGAAGAAACAGAAACAAACAAACAAACAAACAAACAAAC 4764
DB 1586 ACCCTGTCTCGAAGAAACCAAAAAAAAAAAAAAAAAAAAAAAAAAGGCC 1535
```

RESULT 12  
US-08-967-727-5/c

```

; Sequence 5, Application US/08967727
; Patent No. 6025474
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,727
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,365
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-967-727-5

Query Match 1.5%; Score 135.2; DB 3; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCAGTGTGGCAGACGCGCTTTATCCACACTTGGAGGAGCAGAAACAGGCAAGATTTCG 4652
DB 1706 GCAGTGTGGCAGACGCGCTTTATCCACACTTGGAGGAGCAGAAACAGGCAAGATTTCG 1647
QY 4653 AGTTCAAGCCAGCCGCTGTATAGAGTTCACGAGCAGGCGCTACACAGAGAA 4712
DB 1646 AGTTTGAAGCCAGCCGCTGTACAGAGTGTCCAGGACAGGCGATACAAAGAAA 1587
QY 4713 ACCCTGTTTGAAGAAACAGAAACAAACAAACAAACAAACAAACAAAC 4764
DB 1586 ACCCTGTCTCGAAGAAACCAAAAAAAAAAAAAAAAAAAAAAAAAAGGCC 1535
```

RESULT 13

US-08-037-230D-5/C  
Sequence 5, Application US/08037230D

Patent No. 6235525

GENERAL INFORMATION:

APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;  
APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For  
NUMBER OF SEQUENCES: 30  
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felte & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/037,230D

FILING DATE: 26-MARCH-1993

CLASSIFICATION: 435

APPLICATION NUMBER: PCT/US92/04354

FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/807,043

FILING DATE: 12-DECEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,364

FILING DATE: 23-SEPTEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/728,838

FILING DATE: 9-JULY-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/705,702

FILING DATE: 23-MAY-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: LUD 5353

REFERENCE/DOCKET NUMBER: LUD 5353

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 4698 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

US-08-037-230D-5

Query Match

Best Local Similarity 1.5%; Score 135.2; DB 3; Length 4698;

Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCGAGAGCAGGCGATTCTG 4652

DB 1706 GCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCGAGAGCAGGCGATTCTG 1647

QY 4653 AGTTCAAGGCCAGCGCTGTCTATAGAGTGAGTTCCAGAGCAGCCAGGCTACACAGAGAA 4712

DB 1646 AGTTGAGGCCAGCGCTGTCTATAGAGTGAGTGCCAGAGCAGCCAGGCTACAAAGAGAA 1587

QY 4713 ACCCTGTTTGAAGAAACCAAGAAACCAAAACCAAAACCAAAACCAAAAC 4764

DB 1586 ACCCTGCTCGAAAAACCAAAAAACCAAAAAACCAAAAAACCAAAAGGCC 1535

RESULT 14

US-09-583-850-5/C  
Sequence 5, Application US/09583850

Patent No. 6498021

GENERAL INFORMATION:

APPLICANT: Gaugler, Batrice; Van den Eynde, BenoEt;  
APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For  
NUMBER OF SEQUENCES: 30  
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felte & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/583,850

FILING DATE:

APPLICATION NUMBER: 09/583,613

FILING DATE:

APPLICATION NUMBER: PCT/US92/04354

FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/807,043

FILING DATE: 12-DECEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,365

FILING DATE: 23-SEPTEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/728,838

FILING DATE: 9-JULY-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/705,702

FILING DATE: 23-MAY-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: LUD 5353

REFERENCE/DOCKET NUMBER: LUD 5353

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 4698 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

US-09-583-850-5

Query Match

Best Local Similarity 1.5%; Score 135.2; DB 4; Length 4698;

Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCGAGAGCAGGCGATTCTG 4652

DB 1706 GCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCGAGAGCAGGCGATTCTG 1647

QY 4653 AGTTCAAGGCCAGCGCTGTCTATAGAGTGAGTTCCAGAGCAGCCAGGCTACACAGAGAA 4712

DB 1646 AGTTGAGGCCAGCGCTGTCTATAGAGTGAGTGCCAGAGCAGCCAGGCTACAAAGAGAA 1587

QY 4713 ACCCTGTTTGAAGAAACCAAGAAACCAAAACCAAAACCAAAACCAAAAC 4764

DB 1586 ACCCTGCTCGAAAAACCAAAAAACCAAAAAACCAAAAAACCAAAAGGCC 1535

Search completed: September 13, 2003, 19:53:41  
Job time : 441 secs

```

RESULT 15
US-09-579-197-5/C
; Sequence 5, Application US/09579197
; Patent No. 6552180
; GENERAL INFORMATION:
; APPLICANT: Gaugler, Beatrice; Van den Eynde, Benoît;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding
; TITLE OF INVENTION: For Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/579,197
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6552180man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ. ID NO.: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-09-579-197-5

Query Match      1.5%; Score 135.2; DB 4; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCAGTGGTGACACGCCCTTTAATCCAGCACTTGGGAGGACAGACAGGAGCATTTCTG 4652
    ||||| ||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1706 GCAGTCTCTGTGCAAGCTTTAATCCAGCATTTGGAGGACAGACAGGAGCATTTCTG 1647
QY 4653 AGTTCAAGGCCACGCTGGTCTATAGAGTGAAGTCCAGGACAGGAGGCTACACAGAA 4712
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1646 AGTTTGGGCGACCGCTGTCTACAGAGTGTGCCAGGACAGCCAGGAGATCAAAAGAA 1587
QY 4713 ACCCTGTTTGAAGAACAGAAAAACAAAAACAAAAACAAAAACAAAAAC 4764
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1586 ACCCTGTCTCGAAAAACCAAAAAACAAAAACAAAAACAAAAACAAAAAGGCC 1535

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 17:40:33 ; Search time 2039 Seconds  
(without alignments)  
11908.507 Million cell updates/sec

Title: US-09-877-935-1

Sequence: 1 gatctgtgcaaccaagaca.....ctctaggtctgtccaccatg 8995

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

Database :

1: N\_Geneseq\_19Jun03: \*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	8995	100.0	8995	21	AAA48515	Murine villin gene
	2	8969.6	99.7	8993	21	AAA46505	Mouse villin gene
	3	153.2	1.7	23024	22	AAF25499	Nucleotide sequenc
	4	150.2	1.7	21580	24	AAL38337	Complementary str
	5	150.6	1.7	1445	22	AAAC8696	Murine variant 2a1
	6	148.8	1.7	7208	21	AAA40866	Murine tumour necr
	7	145.6	1.6	6789	22	AAF63436	Murine CD33-L4 gen
c	8	145.2	1.6	40116	25	AB226080	Mouse DnaaseX encod

C	9	144.8	1.6	171936	24	ABSS5655	Human SULF2 genom
C	10	144.6	1.6	2221	21	AAA47551	Mouse TANCO 232 co
C	11	144.6	1.6	2221	24	AAD44193	Mouse TANCO 232 CD
C	12	143.2	1.6	17135	25	AAL41384	pgamma6.0FtExbeta-
C	13	142.6	1.6	14707	24	AAL55325	genomic DNA of mou
C	14	142.2	1.6	659158	25	ABX16390	Mouse high growth
C	15	141.8	1.6	15914	22	AAFS5269	Mouse RetL5 genom
C	16	141.8	1.6	21960	23	ABZ72363	Mouse Gene 216 ortt
C	17	141.8	1.6	21960	23	ABX75229	Genomic DNA sequen
C	18	141.4	1.6	1698	20	AAAX2313	Mouse 1-alpha-OHase
C	19	141.4	1.6	4105	25	AAST2316	Mouse 1-alpha-OHase
C	20	140.8	1.6	3815	25	ABSS7008	Genomic DNA encod
C	21	140.8	1.6	20775	24	AAL48783	Murine transcriptit
C	22	140.8	1.6	20775	24	AAL47852	Murine basic helix
C	23	140.8	1.6	20775	24	ABN83212	Mouse transcriptio
C	24	140.4	1.6	3360	19	AAV19608	Mus musculus EPCR
C	25	139.4	1.5	7661	24	ABO88221	Human osteoblast d
C	26	139	1.5	1720	24	ABSS6714	Mouse gene (EMBL
C	27	138.4	1.5	8300	22	AAH25711	Murine DNA polymer
C	28	138.2	1.5	16956	21	AAZ91923	Wild type (C57BL/6
C	29	138	1.5	5973	21	AAZ91924	Wild type (C57BL/6
C	30	138	1.5	11689	21	AAAC9190	Wild type (C57BL/6
C	31	137.8	1.5	13558	22	AAL66272	Mouse plakophilin-
C	32	137.8	1.5	173810	24	ABN83752	Murine apoptase L1
C	33	137.6	1.5	1204	24	ABSS6588	Mouse chromosome 1
C	34	137.4	1.5	1135	24	ABSS6704	Mouse gene (EMBL
C	35	137.4	1.5	1171	24	ABSS6705	Mouse gene (EMBL
C	36	137	1.5	3720	24	AAD27687	Plasmid pBAC84G1 c
C	37	136.8	1.5	2509	19	AAV07515	Vascular endothel
C	38	136.8	1.5	3010	22	AAF86153	Murine JNK activat
C	39	136.2	1.5	24173	24	ABQ78869	Mouse fibroblast g
C	40	136	1.5	3436	24	AAD45148	Human GNG-6 gene f
C	41	136	1.5	20775	24	AAL47852	Murine transcriptit
C	42	136	1.5	20775	24	AAL47853	Murine basic helix
C	43	136	1.5	20775	24	ABN83212	Mouse transcriptid
C	44	135.6	1.5	17004	25	ABV74560	Murine period 2 ge
C	45	135.4	1.5	90050	21	AAZ91925	Wild type (C57BL/6

## ALIGNMENTS

```

RESULT 1
AAA48515
ID    AAA48515 standard; DNA; 8995 BP.

```

AC AAA48515;

DT 19-DEC-2000 (first entry)

Murine villin gene regulatory region.

KW Mouse; villin; regulatory region; digestive tract; colostrum; cancer mouse model; ds

KW colorectal cancer mouse model; ds.

Mus sp. 05

	Key	Location/Qualifiers
FH	misc_signal	3442
FT		

/note= "transcriptional start site"

ET

ET intron

Er

13

misc\_signal

LET

XX	WM200034492-A1.
FN	
XX	
PD	15-JUN-2000.
XX	
PE	09-DEC-1998; 98WO-EP08009.
XX	
XX	09-DEC-1998; 98WO-EP08009.
XX	
PA	(CNRS ) CENT NAT RECH SCI.
XX	(CURRI-) INST CURIE.
XX	
PI	Pinto D, Robine S, Jaisser F, Louvard D;
XX	
DR	WPI; 2000-423433/36.
XX	
PT	Novel nucleotide sequence derived from mouse villin gene for targeted
PT	expression of transgenes in immature and differentiated epithelial
PT	cells of intestine or urogenital tracts
XX	
XX	
PS	Claim 3; Fig 6; 54pp; English.
XX	
CC	The present sequence comprises the murine villin gene regulatory
CC	region, the first Intron and part of the first two exons. It has been
CC	shown that this region directs the expression of the villin gene in the
CC	intestine and uro-genital tracts, and thus could be used in a fusion
CC	gene to direct expression of exogenous genes in these areas. This could
CC	be used, for example, to create a mouse model for colorectal cancer.
XX	
SO	Sequence 8995 BP; 2275 A; 2105 C; 2258 G; 2357 T; 0 other;

Db	541	GAAGACGTGACGGCCTTATTAACACCTGGCACTTATTAACACTTATTAACACTGGCACAGG	600
Qy	601	CGTTACAGSTTTGAAGATCACTTTTCAAAACACAGAAAGATGCTGCTGCTTACG	660
Db	601	CGTTACAGSTTTGAAGATCACTTTTCAAAACACAGAAAGATGCTGCTGCTTACG	660
Qy	661	GTACGAGCACTGGCTGCAGAAAGATATTTAGTGAAGCTACCTTACAAATATCTTT	720
Db	661	GTACGAGCACTGGCTGCAGAAAGATATTTAGTGAAGCTACCTTACAAATATCTTT	720
Qy	721	GCACCTATACATACAGCTGTCCAATGTGCTAACTCCTAGTCCACAGATGGCTTTACA	780
Db	721	GCACCTATACATACAGCTGTCCAATGTGCTAACTCCTAGTCCACAGATGGCTTTACA	780
Qy	781	CTCGTTTCGTGTTTCCCATCTGTTGATTTGTCCAGAACCAAAATTAGAAATGTGGGT	840
Db	781	CTCGTTTCGTGTTTCCCATCTGTTGATTTGTCCAGAACCAAAATTAGAAATGTGGGT	840
Qy	841	ATTTATTTGTGTGCTGAGAGACACATCCAGAGGCTTTTACATTTTCAGGACATGGTTTAC	900
Db	841	ATTTATTTGTGTGCTGAGAGACACATCCAGAGGCTTTTACATTTTCAGGACATGGTTTAC	900
Qy	901	TAACTGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTTATATTTACTTATTTTGTGT	960
Db	901	TAACTGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTTATATTTACTTATTTTGTGT	960
Qy	961	GCATGAGGTAGGCATGTATACGTATGTATAGAGTCAATGCATGTGGCTGTACCTCAAA	1020
Db	961	GCATGAGGTAGGCATGTATACGTATGTATAGAGTCAATGCATGTGGCTGTACCTCAAA	1020

Query Match	100.0%;	Score 8995;	DB 21;	Length 8995;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 8995;	Conservative	0;	Mismatches	0;
			Indels	Gaps
QY	1	GATCGTGACACCAAGGACACATGTGTGCCAGCACTGGGGAGGTGAGGAGGAGGTC	60	
DB	1	GATCGTGACACCAAGGACACATGTGTGCCAGCACTGGGGAGGTGAGGAGGAGGTC	60	
QY	61	GAAGTTTAAGTCATCCTTGGTTACATAGCAAGGTTTCAGCCAGCTTCAGCTACATGAAA	120	
DB	61	GAAGTTTAAGTCATCCTTGGTTACATAGCAAGGTTTCAGCCAGCTTCAGCTACATGAAA	120	
QY	121	CCTTTGTTGTTGTTGTTGTTGTTTAAAGCACTTAATTAATACCATAGAGAGTTGG	180	
DB	121	CCTTTGTTGTTGTTGTTGTTGTTTAAAGCACTTAATTAATTAATACCATAGAGAGTTGG	180	
QY	181	CAGTGTGTGGCAGACACCTTAAATTCACATATTCAGAGGCGAGACAGGACAGATCTCT	240	
DB	181	CAGTGTGTGGCAGACACCTTAAATTCACATATTCAGAGGCGAGACAGGACAGATCTCT	240	
QY	241	GAGTTCGAAGTCAGCCTAGTGTGCAGAACTAGTTCCAGATGGCAGAGGCTCACAGAGA	300	
DB	241	GAGTTCGAAGTCAGCCTAGTGTGCAGAACTAGTTCCAGATGGCAGAGGCTCACAGAGA	300	
QY	301	AACCTGTCTCATAAACCAAGTAGTAGTAGTATAGTATGCCATAGAGAAAATTGGA	360	
DB	301	AACCTGTCTCATAAACCAAGTAGTAGTAGTATAGTATGCCATAGAGAAAATTGGA	360	
QY	361	GTCCATTCAGATGAGCACCCTATTAATGATGTTCTTGACCCAGTAGACTAATGTCA	420	
DB	361	GTCCATTCAGATGAGCACCCTATTAATGATGTTCTTGACCCAGTAGACTAATGTCA	420	
QY	421	TGGGAAAAGGGGATGCGACTGTCCTAGATTAAAAAGTGCTGAGGCGATGCTATTCTCA	480	
DB	421	TGGGAAAAGGGGATGCGACTGTCCTAGATTAAAAAGTGCTGAGGCGATGCTATTCTCA	480	
QY	481	TTTGATTCATATGAAAAAGGCTGATAGAGCCCAAGAGTAGTGAAGTGGAGCTCTGGACT	540	
DB	481	TTTGATTCATATGAAAAAGGCTGATAGAGCCCAAGAGTAGTGAAGTGGAGCTCTGGACT	540	
QY	541	GAGACGTGACGCGCTTTAAACACTGGCAGCTTATTAACACTTATTAACACTGGCAGG	600	
DB	541	GAGACGTGACGCGCTTTAAACACTGGCAGCTTATTAACACTTATTAACACTGGCAGG	600	

[illegible]

OY	1661	TTAGCATGTGCTGTGAGAGATAGAAAAATGAAACATTGAGTCTAGTCTGGAAACCA	1740
Db	1661	TTAGCATGTGCTGTGAGAGATAGAAAAATGAAACATTGAGTCTAGTCTGGAAACCA	1740
OY	1741	GAGGAGGCGAAGAACCCACCTCCGAAAGTCTCTGAGGCTTCACATACACCTTCAT	1800
Db	1741	GAGGAGGCGAAGAACCCACCTCCGAAAGTCTCTGAGGCTTCACATACACCTTCAT	1800
OY	1801	AATAGTTTCAATGATTAATTAATTAATTTAGTAATCTTTTAAAGGTATATGTTGGAGGA	1860
Db	1801	AATAGTTTCAATGATTAATTAATTAATTTAGTAATCTTTTAAAGGTATATGTTGGAGGA	1860
OY	1861	GAGATGGCTCAGCTTCCAGAGACCTTGCTCTTGAGAGACCTTAGATTCACTTCCC	1920
Db	1861	GAGATGGCTCAGCTTCCAGAGACCTTGCTCTTGAGAGACCTTAGATTCACTTCCC	1920
OY	1921	AGGACTCATATGGTGGCTCACAGCATCTGTAAATCCAGTCCAGAGGTTCCACACCT	1980
Db	1921	AGGACTCATATGGTGGCTCACAGCATCTGTAAATCCAGTCCAGAGGTTCCACACCT	1980
OY	1981	CTTCTGGCCTCCACAGGCACCACTACATATAGTACACAGACATACATGACAGCAAAAC	2040
Db	1981	CTTCTGGCCTCCACAGGCACCACTACATATAGTACACAGACATACATGACAGCAAAAC	2040
OY	2041	CATACACATTAATTAATTAAGAACTTTAAAGGTGCATGTGTTGTAACATTGTCT	2100
Db	2041	CATACACATTAATTAATTAAGAACTTTAAAGGTGCATGTGTTGTAACATTGTCT	2100
OY	2101	TACACATCTCATTTGAAGACATGTACAAAGCACACACTGAAAGAGGATCTGGGCTGG	2160
Db	2101	TACACATCTCATTTGAAGACATGTACAAAGCACACACTGAAAGAGGATCTGGGCTGG	2160
OY	2161	AGATGGCTCAGCGTTAAAGACACTGACTGCTCTCCGAAGAGAGTCTGAGTTCAAT	2220
Db	2161	AGATGGCTCAGCGTTAAAGACACTGACTGCTCTCCGAAGAGAGTCTGAGTTCAAT	2220
OY	2221	CCTGACACACCATGTGTGCTCACACACATCCATTAATGATCTGACACCTCTTGTGT	2280
Db	2221	CCTGACACACCATGTGTGCTCACACACATCCATTAATGATCTGACACCTCTTGTGT	2280
OY	2281	GCATCTGAGACAGCTGCAGAGCTACAGTGTACTTGAATACATTAATTAATCTTTT	2340
Db	2281	GCATCTGAGACAGCTGCAGAGCTACAGTGTACTTGAATACATTAATTAATCTTTT	2340
OY	2341	TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTAATGACAGTGCACG	2400
Db	2341	TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTAATGACAGTGCACG	2400
OY	2401	GGTATATATCTATCCTGGAGTTTTTCCCTTCCGCTTGGCTTGAACAAACAGCACTGGAGT	2460
Db	2401	GGTATATATCTATCCTGGAGTTTTTCCCTTCCGCTTGGCTTGAACAAACAGCACTGGAGT	2460
OY	2461	CCCCCTTTTCATTCACAAAGAGGGGTGCATTTATTTCTGAAACAAACAGCACTGGAGT	2520
Db	2461	CCCCCTTTTCATTCACAAAGAGGGGTGCATTTATTTCTGAAACAAACAGCACTGGAGT	2520
OY	2521	ATGTTTACTGTCTTGTCTGACTATAGACAGCGCAGCGCGCGCGCAACACACACAC	2580
Db	2521	ATGTTTACTGTCTTGTCTGACTATAGACAGCGCAGCGCGCGCGCGCAACACACACAC	2580
OY	2581	ACACACACACACACACACACACACACACATTCATGCTCCAGAGCTCTTGGGAAGTCA	2640
Db	2581	ACACACACACACACACACACACACACACATTCATGCTCCAGAGCTCTTGGGAAGTCA	2640
OY	2641	AGAAAGAGGCTGCCCTCAAAACAGATCTTTCATCTTCCCTCTTAAAGAGACACAGATTCC	2700
Db	2641	AGAAAGAGGCTGCCCTCAAAACAGATCTTTCATCTTCCCTCTTAAAGAGACACAGATTCC	2700
OY	2701	AAGGTGGCAGAAATCTTACAGGGGGCGAGGCGAGGAGGGGGAAGAGCCATGTGTTCC	2760
Db	2701	AAGGTGGCAGAAATCTTACAGGGGGCGAGGCGAGGAGGGGGAAGAGCCATGTGTTCC	2760

QY	2761	AGAGACCTACACAGCAGAGGGGACAGCAAGGCAATGCCAGTCCAGGGGACGGGAGGTGGAGG	2820
Db	2761	AGAGACCTACACAGCAGAGGGGACAGCAAGGCAAGATCCCCAGTCCAGGGGACGGGAGGTGGAGG	2820
QY	2821	CCCTGTTCCGAGGGAAGGAGGAGGCGGGGAGAAACAGGTTCAAAGGACACAGGTTTATGGCA	2880
Db	2821	CCCTTGTTCGAGGGAAGGAGGAGGCGGGGAGAAACAGGTTCAAAGGACACAGGTTTATGGCA	2880
QY	2881	GCTCATAAAGATGAGAGTCTGTGGCTCAGTCAAGAAAGGAGGAAGAGGGAAGGAGGCCCTTGT	2940
Db	2881	GOTCATAAAGATGAGAGTCTGTGGCTCAGTCAAGAAAGGAGGAAGAGGGAAGGAGGCCCTTGT	2940
QY	2941	GCCCACTAAGGAGGGGTATGTCTGAGTACGAGAGATCTGCAGAGGCTGCCAGAGCCCAAC	3000
Db	2941	GCCCACTAAGGAGGGGTATGTCTGAGTACGAGAGATCTGCAGAGGCTGCCAGAGCCCAAC	3000
QY	3001	CTGTCTGTCCAAAGGGAACCCCAAGTGAACCTCGGGCTTGGGGGTCTGAGTCCAGCTA	3060
Db	3001	CTGTCTGTCCAAAGGGAACCCCAAGTGAACCTCGGGCTTGGGGGTCTGAGTCCAGCTA	3060
QY	3061	CAAGACCCCAAGAGTCTACTCTCAATCCCCATCCAGTGGCCCCCTCGCCCCACACCCCA	3120
Db	3061	CAAGACCCCAAGAGTCTACTCTCAATCCCCATCCAGTGGCCCCCTCGCCCCACACCCCA	3120
QY	3121	CCCCGACATCCCGTCCACTTCTTAGGGGTGAGAGGATCTGAGGGGCTG	3180
Db	3121	CCCCGACATCCCGTCCACTTCTTAGGGGTGAGAGGATCTGAGGGGCTGAGGGGCTG	3180
QY	3181	CTACCTGAGGTAAAGGCCAGGTCTTACCCGGAAGTGAACCCCATCCCTGAAGCTGCAGA	3240
Db	3181	CTACCTGAGGTAAAGGCCAGGTCTTACCCGGAAGTGAACCCCATCCCTGAAGCTGCAGA	3240
QY	3241	GCCAAAGGCGGGGGACACAGGAGCTCAGAGCTGTGAGGCTGTGAGGCTCTAGGTTCC	3300
Db	3241	GCCAAAGGCGGGGGACACAGGAGCTCAGAGCTGTGAGGCTGTGAGGCTCTAGGTTCC	3300
QY	3301	AGGGACCTGGACACTTCTCCACCCCCCATCATTCTCTGTGGGCCCTATCTTCC	3360
Db	3301	AGGGACCTGGACACTTCTCCACCCCCCATCATTCTCTGTGGGCCCTATCTTCC	3360
QY	3361	CTTATATGCTAAAGGAATCTCTGGGGGGGGGGTGGTGAAGGAATAAGTTCGTTCG	3420
Db	3361	CTTATATGCTAAAGGAATCTCTGGGGGGGGGGTGGTGAAGGAATAAGTTCGTTCG	3420
QY	3421	GTCCTCTGACAGCCAGCTTGCCACAACCTCTCAAGATCTCCAGAGGTGTGGCTGCTTTC	3480
Db	3421	GTCCTCTGACAGCCAGCTTGCCACAACCTCTCAAGATCTCCAGAGGTGTGGCTGCTTTC	3480
QY	3481	CAGACAGGTAAGGCAATGGGTGGGGACACATGTGTGACACAGAGGTGTGAGGGGACAG	3540
Db	3481	CAGACAGGTAAGGCAATGGGTGGGGACACATGTGTGACACAGAGGTGTGAGGGGACAG	3540
QY	3541	GGTCTCTGCTCTCTCTGGGACCGTGTCTTCTGTACACTTGGTATAGGTTGGGGG	3600
Db	3541	GGTCTCTGCTCTCTCTGGGACCGTGTCTTCTGTACACTTGGTATAGGTTGGGGG	3600
QY	3601	TGAGGTAGGTGCTCTGAAACTCTGAAAGAAAGCAAGAGCCAGCAGCTGTGGGCT	3660
Db	3601	TGAGGTAGGTGCTCTGAAACTCTGAAAGAAAGCAAGAGCCAGCAGCTGTGGGCT	3660
QY	3661	TCAATGAAGAAATTACAGACAGCCCTTCTGTAAAGTACACTTTCGCTTATCTGTGTAG	3720
Db	3661	TCAATGAAGAAATTACAGACAGCCCTTCTGTAAAGTACACTTTCGCTTATCTGTGTAG	3720
QY	3721	ATTTCCTTGGGACCAAGGTGGCTCTGGGACCTCAGATTTCTACATTAATAACAGACAGT	3780
Db	3721	ATTTCCTTGGGACCAAGGTGGCTCTGGGACCTCAGATTTCTACATTAATAACAGACAGT	3780
QY	3781	CCTGAGACTTGGACCTCGCTGCTGATTTACTTCTCTGTGGCTGCTATTTCTGTGT	3840
Db	3781	CCTGAGACTTGGACCTCGCTGCTGATTTACTTCTCTGTGGCTGCTATTTCTGTGT	3840
QY	3841	TCATGTCTTACACATCTGAAGATGTTCTTTTGTGTACACATTCCTGTACACTCTGGGA	3900

|||||  
Db 3841 TCATGCTTACACATCTGAAATGTTTCTTTGTGTACACATTCCTCCCTGACACTCTGGGA 3900  
QY 3901 GGTGCTATCCTTGGCAGATGTATCTGGATGTAACTGCAGCCAGCAGAGAGAGGG 3960  
Db 3901 GGTGCTATCCTTGGCAGATGTATCTGGATGTAACTGCAGCCAGCAGAGAGAGGG 3960  
QY 3961 AGAGTCAGAGAGCTGTGTCTTAGGCCCTATTAGGCCCTGGACATCAACCCCTTCTAGAAAT 4020  
Db 3961 AGAGTCAGAGAGCTGTGTCTTAGGCCCTATTAGGCCCTGGACATCAACCCCTTCTAGAAAT 4020  
QY 4021 GGCCCTCCATTTTGGTGTACCATGTATTTTATATCAGATGGCCAGTAAAGCCA 4080  
Db 4021 GGCCCTCCATTTTGGTGTACCATGTATTTTATATCAGATGGCCAGTAAAGCCA 4080  
QY 4081 AACCTGCCAGAAAGTTTGGGACTCAGTCAGACCAAGGTTATCTGCTCAGAAATCCCTG 4140  
Db 4081 AACCTGCCAGAAAGTTTGGGACTCAGTCAGACCAAGGTTATCTGCTCAGAAATCCCTG 4140  
QY 4141 TCACCTGAGGTTGGGAGAAATCTGCTGGGGCTTCAGGCTTGGTTAGCAGAGAGGT 4200  
Db 4141 TCACCTGAGGTTGGGAGAAATCTGCTGGGGCTTCAGGCTTGGTTAGCAGAGAGGT 4200  
QY 4201 ATCTTGTATAGGGAATGACATGTATGTGTACTATTCATTCCTGCTCAATTAAG 4260  
Db 4201 ATCTTGTATAGGGAATGACATGTATGTGTACTATTCATTCCTGCTCAATTAAG 4260  
QY 4261 CTGGAACATAAAACCCAGCGAGCCAGGATCTCTACAGTTGTACCCCAAGAACCA 4320  
Db 4261 CTGGAACATAAAACCCAGCGAGCCAGGATCTCTACAGTTGTACCCCAAGAACCA 4320  
QY 4321 AGCAGTAGATATGCAAGATAGTAGCTGGGAGAAAGAACTTAACCCCCCAAG 4380  
Db 4321 AGCAGTAGATATGCAAGATAGTAGCTGGGAGAAAGAACTTAACCCCCCAAG 4380  
QY 4381 GCCACAGGTTCCGTTCCCTAGTTTCAATAGCCAGTATGAGTGTAGCTATAGGCTG 4440  
Db 4381 GCCACAGGTTCCGTTCCCTAGTTTCAATAGCCAGTATGAGTGTAGCTATAGGCTG 4440  
QY 4441 TGAGTTGATAGTACAAAGATGATGTATCTGATGTGTATGTATATCTGACAC 4500  
Db 4441 TGAGTTGATAGTACAAAGATGATGTATCTGATGTGTATGTATATCTGACAC 4500  
QY 4501 TTGGGAGGCTGAAGCAGAGAGATTGCTATATGTTTGAAGCCAGCTGAGTATAGAGCA 4560  
Db 4501 TTGGGAGGCTGAAGCAGAGAGATTGCTATATGTTTGAAGCCAGCTGAGTATAGAGCA 4560  
QY 4561 GACTTGTCTTTAAGAAAAAATGAAGAGCCAGCAGAGTGGGACACGCCCTTAATCCCA 4620  
Db 4561 GACTTGTCTTTAAGAAAAAATGAAGAGCCAGCAGAGTGGGACACGCCCTTAATCCCA 4620  
QY 4621 GCACCTTGGAGGAGAGAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGTCTATAGAGT 4680  
Db 4621 GCACCTTGGAGGAGAGAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGTCTATAGAGT 4680  
QY 4681 GAGTTCAGAGACGCCAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAGAAACAA 4740  
Db 4681 GAGTTCAGAGACGCCAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAGAAACAA 4740  
QY 4741 AACAAAAACAACAAAAACAACCAACCAACCAACCAACCTCTCATCTCTCTC 4800  
Db 4741 AACAAAAACAACAAAAACAACCAACCAACCAACCAACCTCTCATCTCTCTC 4800  
QY 4801 TAGGCTGTCTGTAGGTGTAGATTGGGACTTCAGACTTATATTAATAGGCC 4860  
Db 4801 TAGGCTGTCTGTAGGTGTAGATTGGGACTTCAGACTTATATTAATAGGCC 4860  
QY 4861 TTTTATACAGTGTAGAGACGAGAAAGTTTCTAGTCTGGACACAGTGGACCTTGAGA 4920  
Db 4861 TTTTATACAGTGTGTAGAGACGAGAAAGTTTCTAGTCTGGACACAGTGGACCTTGAGA 4920  
QY 4921 AACTACTCTTGGACCCCAAAATTTCTGGGAAGCTTCTGAGAGAGTGTCTCCGAT 4980  
|||||

Db 4921 AACTACTCTTGGACCCCAAAATTTCTGGGAAGCTTCTGAGAGAGTGTCTCCGAT 4980  
QY 4981 CAGACTACTGTGTAGAGGACAGAGAGAGGTTGGAGAAATTTGGTGGACAGACTT 5040  
Db 4981 CAGACTACTGTGTAGAGGACAGAGAGAGGTTGGAGAAATTTGGTGGACAGACTT 5040  
QY 5041 GGAACAGAGACAGAGAGGGGAGGCATCCAGATTTCTGAACATGTAGCTGACTTTTGGT 5100  
Db 5041 GGAACAGAGACAGAGAGGGGAGGCATCCAAATTTCTGAACATGTAGCTGACTTTTGGT 5100  
QY 5101 TCTGTGGGTGACAAAGTGTCCCGAGGATGGGCTGTAGAAAGGGGACCGGGGTGAGCC 5160  
Db 5101 TCTGTGGGTGACAAAGTGTCCCGAGGATGGGCTGTAGAAAGGGGACCGGGGTGAGCC 5160  
QY 5161 AATGAGTTCAAGTTGAGGGAACATCCAGCCAGGAGCTCTTGTGCGCAACCTTAAGATG 5220  
Db 5161 AATGAGTTCAAGTTGAGGGAACATCCAGCCAGGAGCTCTTGTGCGCAACCTTAAGATG 5220  
QY 5221 AGAGCCCTTAACCTTCCCTGAAAGTTTGGGAGACAGAGAGCTGAGAGATCTTCTA 5280  
Db 5221 AGAGCCCTTAACCTTCCCTGAAAGTTTGGGAGACAGAGAGCTGAGAGATCTTCTA 5280  
QY 5281 GGGTGAAGGAGGATCTGCTGACCAACATGGCTAGAGGAGAGAAAGCAGTTGGACAG 5340  
Db 5281 GGGTGAAGGAGGATCTGCTGACCAACATGGCTAGAGGAGAGAAAGCAGTTGGACAG 5340  
QY 5341 TTACCCCTCAGAACCAAGCCATCCCTTGGCTCTAAGAGAGGCTGGGCCCTTCTGTTT 5400  
Db 5341 TTACCCCTCAGAACCAAGCCATCCCTTGGCTCTAAGAGAGGCTGGGCCCTTCTGTTT 5400  
QY 5401 AAGAATCTACTTTTCTCAGAGAGAGGACAGAGCTTTGTCCCTCCCTGTTGTCAA 5460  
Db 5401 AAGAATCTACTTTTCTCAGAGAGAGGACAGAGCTTTGTCCCTCCCTGTTGTCAA 5460  
QY 5461 TAAACACCCCTGTGTATTAATTAATTTACTCTGAGTTTGTCTCCAGAGAGTCCA 5520  
Db 5461 TAAACACCCCTGTGTATTAATTAATTTACTCTGAGTTTGTCTCCAGAGAGTCCA 5520  
QY 5521 TCTGTAGAGCTGCTGCTTAACCTACCAAGATGATGGCCCAATTCCTCAGAGAGAG 5580  
Db 5521 TCTGTAGAGCTGCTGCTTAACCTACCAAGATGATGGCCCAATTCCTCAGAGAGAG 5580  
QY 5581 TGCAGAGAGAGGCTTGAAGAAAGGTTAACAGTAAACAATATGGCCAGATTAACACAAA 5640  
Db 5581 TGCAGAGAGAGGCTTGAAGAAAGGTTAACAGTAAACAATATGGCCAGATTAACACAAA 5640  
QY 5641 ACTACTATCTTTGTACCAAAATTTGTTTGTGCAACAGAGAGGGGTGTGAGTGTAT 5700  
Db 5641 ACTACTATCTTTGTACCAAAATTTGTTTGTGCAACAGAGAGGGGTGTGAGTGTAT 5700  
QY 5701 GT 5760  
Db 5701 GT 5760  
QY 5761 CTGGGGGACTTTTCAAGCTTAAGAAATATCTGATATTTGGGCCCATGCCAAGGGTAT 5820  
Db 5761 CTGGGGGACTTTTCAAGCTTAAGAAATATCTGATATTTGGGCCCATGCCAAGGGTAT 5820  
QY 5821 TGGGAGAGTCAAGGCTTCTGCAAAACACAGTAACTGCCCAAGATGATTTGGTGGCTGAA 5880  
Db 5821 TGGGAGAGTCAAGGCTTCTGCAAAACACAGTAACTGCCCAAGATGATTTGGTGGCTGAA 5880  
QY 5881 TCACCAAGGGGAGGCTGATCAGAGTGAGAGAAATCACAAGATTAAGCCACCTGTGGG 5940  
Db 5881 TCACCAAGGGGAGGCTGATCAGAGTGAGAGAAATCACAAGATTAAGCCACCTGTGGG 5940  
QY 5941 GCTCAGAGAGGAGGTTTACAAAGGTTAAAGGCCAAGCCATTTATATCCAAACAAGTAC 6000  
Db 5941 GCTCAGAGAGGAGGTTTACAAAGGTTAAAGGCCAAGCCATTTATATCCAAACAAGTAC 6000  
QY 6001 TCAAAATCAAAAGTCAAGGAGATTTAGCTGGAGAGATGGGGCTGTCAGTGTGGAGACC 6060  
Db 6001 TCAAAATCAAAAGTCAAGGAGATTTAGCTGGAGAGATGGGGCTGTCAGTGTGGAGACC 6060

OY	6061	TGACCTTGCACTTATTAAGTACTACGTAGGCCAAGAGCAGTCACAGAGGGGTACTGGGTCCCTA	6120
Dd	6061	TGACCTTGCACTTATTAAGTACTACGTAGGCCAAGAGCAGTCACAGAGGGGTACTGGGTCCCTA	6120
OY	6121	CTCAGCTTTGGAGCGACGCAGTGGGAAGAATGGGTGACACCTCATTCGATGGAAGAGGGCTGAG	6180
Dd	6121	CTCAGCTTTGGAGCGACGCAGTGGGAAGAATGGGTGACACCTCATTCGATGGAAGAGGGCTGAG	6180
OY	6181	CACCACGAGTACAAGTGTTCCTCTGTCTCTATCCAGAGATTCCTGGCCACTTTTCCAAG	6240
Dd	6181	CACCACGAGTACAAGTGTTCCTCTGTCTCTATCCAGAGATTCCTGGCCACTTTTCCAAG	6240
OY	6241	GACTTAGGACTCATCTCTGTGTGGAAACAAGATATCCAAAGCCCTTAAGCCCCATTGCTCT	6300
Dd	6241	GACTTAGGACTCATCTCTGTGTGGAAACAAGATATCCAAAGCCCTTAAGCCCCATTGCTCT	6300
OY	6301	AATTAAATACAACCCCCTGGGGGATGACAGGCTCTGAGCAGCACAGAGCTTTTAAAAAGTCC	6360
Dd	6301	AATTAAATACAACCCCCTGGGGGATGACAGGCTCTGAGCAGCACAGAGCTTTTAAAAAGTCC	6360
OY	6361	CCAGGTGATTTGTATCAGCAGCTGGAAACAACACAGCTACAGTTTCAAACGAAAAGAGGC	6420
Dd	6361	CCAGGTGATTTGTATCAGCAGCTGGAAACAACACAGCTACAGTTTCAAACGAAAAGAGGC	6420
OY	6421	AAAGCTAGGAAAGGTTGGATGGGAGGACCTTCTCCAGGCGACATAGATGGAGGCTGGTT	6480
Dd	6421	AAAGCTAGGAAAGGTTGGATGGGAGGACCTTCTCCAGGCGACATAGATGGAGGCTGGTT	6480
OY	6481	AGCAGTGTGGCAGCTTCTCTCTGCTCATATAGCTATAGCTATCCATCCACTCATCCATCAT	6540
Dd	6481	AGCAGTGTGGCAGCTTCTCTCTGCTCATATAGCTATAGCTATCCATCCACTCATCCATCAT	6540
OY	6541	ACACCCACCCATCCATTTATGACACCCATCCCTTCATCCATCCATCTATCCAGCTAACCCAC	6600
Dd	6541	ACACCCACCCATCCATTTATGACACCCATCCCTTCATCCATCCATCTATCCAGCTAACCCAC	6600
OY	6601	CCAGGCATCATTCACAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCATYT	6660
Dd	6601	CCAGGCATCATTCACAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCATYT	6660
OY	6661	CATTATTCACAACAGAACTGGTATTTGACTAAATGGGAGATTTAATTAATTTTTAGA	6720
Dd	6661	CATTATTCACAACAGAACTGGTATTTGACTAAATGGGAGATTTAATTAATTTTTAGA	6720
OY	6721	AGCTCTGTGATGACTGATGTGCATGATATGTGAGACAGTACATPACCACGACGACAGTGTG	6780
Dd	6721	AGCTCTGTGATGACTGATGTGCATGATATGTGAGACAGTACATPACCACGACGACAGTGTG	6780
OY	6781	TGGCAATCGAAGAAAGTTTGGGTGTGTTTCTCTCCACCGCTGGGGTCTGGGGA	6840
Dd	6781	TGGCAATCGAAGAAAGTTTGGGTGTGTTTCTCTCTCCACCGCTGGGGTCTGGGGA	6840
OY	6841	TTGAACATCAATTTATCGGGCTGGTGGGAGAGTCTTAAACACGACGAGCATTTTCTGTACA	6900
Dd	6841	TTGAACATCAATTTATCGGGCTGGTGGGAGAGTCTTAAACACGACGAGCATTTTCTGTACA	6900
OY	6901	CATCATTAATTTATGAAAAGCATCTTATGTAGTCCAGGCTGGCTCAAGCTTGATGTGCG	6960
Dd	6901	CATCATTAATTTATGAAAAGCATCTTATGTAGTCCAGGCTGGCTCAAGCTTGATGTGCG	6960
OY	6961	CCAGGAGTACCTTTAACTCCTGCTCTTCCAGCCTCCACCCGAGTGGTATGATTTACAGGT	7020
Dd	6961	CCAGGAGTACCTTTAACTCCTGCTCTTCCAGCCTCCACCCGAGTGGTATGATTTACAGGT	7020
OY	7021	GTTCAACATGAGTAATGCTTTAATCCAGCACTGTGTGGGGGGGGGGGAGGAGCGATC	7080
Dd	7021	GTTCAACATGAGTAATGCTTTAATCCAGCACTGTGTGGGGGGGGGGGAGGAGCGATC	7080
OY	7081	CCTGAGTTGGAGGCCAGTTTGGTCTACAGAGTTTCAGGATACCTGGGGCTATACAGGGAA	7140
Dd	7081	CCTGAGTTGGAGGCCAGTTTGGTCTACAGAGTTTCAGGATACCTGGGGCTATACAGGGAA	7140

OY	7141	ACCCTATCCCAACAAACAACAAACAACAAAAAATTTCTGTGTCATATTATCAGAGA	7200
Dd	7141	ACCCATATCCCAACAAACAACAAACAACAAAAATTTCTGTGTCATATTATCAGAGA	7200
OY	7201	TTAGGGATTTTAGAGGGTGTAGGGGTGGGGGGGAGGAGCTCATGCTTCTTTTGATTT	7260
Dd	7201	TTAGGGATTTTAGAGGGTGTAGGGGTGGGGGGGAGGAGCTCATGCTTCTTTTGATTT	7260
OY	7261	ATAATAGTAAGTACTCTCAGAGATGCATATCTATCTATCTATCTATCTATCTATCT	7320
Dd	7261	ATAATAGTAAGTACTCTCAGAGATGCATATCTATCTATCTATCTATCTATCTATCT	7320
OY	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	7380
Dd	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	7380
OY	7381	CTGCTTTGACCTGAATGCTGCTATTTCTGGGTCAACCTTCACCCCAGTGGGGTTT	7440
Dd	7381	CTGCTTTGACCTGAATGCTGCTATTTCTGGGTCAACCTTCACCCCAGTGGGGTTT	7440
OY	7441	ACCAACACCCAGACATTTATTTTATTTTGTGTTTTATTTATTTAATCTAGAGCTCAGGTTG	7500
Dd	7441	ACCAACACCCAGACATTTATTTTATTTTGTGTTTTATTTATTTAATCTAGAGCTCAGGTTG	7500
OY	7501	GGACTCAGGGTCTTGCGATGCTAAGCAAGCTCTCTGCCACAGATGTCAGCTCCAGTCC	7560
Dd	7501	GGACTCAGGGTCTTGCGATGCTAAGCAAGCTCTCTGCCACAGATGTCAGCTCCAGTCC	7560
OY	7561	CCATTTTGTGTAGGTGACTGTGTGACAGTTGTCTATTTCCGAGGGCTATGTAGCTCTCTC	7620
Dd	7561	CCATTTTGTGTAGGTGACTGTGTGACAGTTGTCTATTTCCGAGGGCTATGTAGGCTCTCTC	7620
OY	7621	CACCTCCAGTTCAGCACTTTCTGTGATCCCCAGTGGGGGCAACTCTGTGCTCACCA	7680
Dd	7621	CACCTCCAGTTCAGCACTTTCTGTGATCCCCAGTGGGGGCAACTCTGTGCTCACCA	7680
OY	7681	GTCGCCGTGTCCTGTCTCTAGACCTACATATTTGCCCTGTCTGTAACAGTTCAATGTAATG	7740
Dd	7681	GTCGCCGTGTCCTGTCTCTAGACCTACATATTTGCCCTGTCTGTAACAGTTCAATGTAATG	7740
OY	7741	GGATGCGTTCCTGTGTATCTTTTATGTGCTGGCCCCCTTATCTTAGCAGACTTTGTGTG	7800
Dd	7741	GGATGCGTTCCTGTGTATCTTTTATGTGCTGGCCCCCTTATCTTAGCAGACTTTGTGTG	7800
OY	7801	GGCCATGCTGTAGCTACTACTCTATCTATCTATCATCATCTTATGTGCTTATATGTCTCTT	7860
Dd	7801	GGCCATGCTGTAGCTACTACTCTATCTATCTATCATCATCTTATGTGCTTATATGTCTCTT	7860
OY	7861	TGTGTGATTAACCACTTCTGTTCATTTACTGATGGAATTTGTGGCCCCACCCCCAC	7920
Dd	7861	TGTGTGATTAACCACTTCTGTTCATTTACTGATGGAATTTGTGGCCCCACCCCCAC	7920
OY	7921	CCTTTTTTTTTTATTTGAGACAAGGCTTCTGTGTATCTTGTGCAATCTTGGCGTCTCT	7980
Dd	7921	CCTTTTTTTTTTATTTGAGACAAGGCTTCTGTGTATCTTGTGCAATCTTGGCGTCTCT	7980
OY	7981	GAGCTCACTGTGTAGACAGGCTGTGAGAGCTGTGCTCCACTTTTGAACACTCTGTGAAC	8040
Dd	7981	GAGCTCACTGTGTAGACAGGCTGTGAGAGCTGTGCTCCACTTTTGAACACTCTGTGAAC	8040
OY	8041	AGAGTAGCCATGAACCTTCCAAGACAATTTCTGTGTTTTGTTTTTACATTTGTGTG	8100
Dd	8041	AGAGTAGCCATGAACCTTCCAAGACAATTTCTGTGTTTTGTTTTTACATTTGTGTG	8100
OY	8101	TGTATGCGTGTATATGTGCAATGTTTGTGTCTTCAAGGTGCTACATGTGTACTGTGTG	8160
Dd	8101	TGTATGCGTGTATATGTGCAATGTTTGTGTCTTCAAGGTGCTACATGTGTACTGTGTG	8160
OY	8161	TGGGACAGAGAACAAACCGATGTGSCATTTCTCGATGATCTACGAGATCTGTGTAATATGTA	8220
Dd	8161	TGGGACAGAGAACAAACCGATGTGSCATTTCTCGATGATCTACGAGATCTGTGTAATATGTA	8220
OY	8221	TGTTTATATGTATGTTTTATTTATGTGTGCCCAAGTATGAGGATTTTGTGTGAGTTTTTAC	8280



QY 421 TGGGGAAGGGGATGGAGCTGCTCAGATTAAAAAGTGTGAGGGGATGCTATTCTCAA 480  
|||||  
Db 421 TGGGGAAGGGGATGGAGCTGCTCAGATTAAAAAGTGTGAGGGGATGCTATTCTCAA 480  
QY 481 TTTGATTCCATTGAAAAAGGCTGATPAGGCCCAAGAGAAATGGAATGSGGACTGTGACT 540  
|||||  
Db 481 TTTGATTCCATTGAAAAAGGCTGATPAGGCCCAAGAGAAATGGAATGSGGACTGTGACT 540  
QY 541 GAAGAGCTGAGGGCCCTTATTAACACTGGGACTTATTAACACTTATTAACACTGGGACAG 600  
|||||  
Db 541 GAAGAGCTGAGGGCCCTTATTAACACTGGGACTTATTAACACTTATTAACACTGGGACAG 600  
QY 601 CGTTCAAGTTTGAAGATCACTTTCAAAACCCACAGAAAGTGTGCTGCTGCTCAGC 660  
|||||  
Db 601 CGTTCAAGTTTGAAGATCACTTTCAAAACCCACAGAAAGTGTGCTGCTGCTCAGC 660  
QY 661 GTAGGAGACACTGGCTGCAGAAAGATGATATTAGTGAAGCTTACCTTACAAATCTTT 720  
|||||  
Db 661 GTAGGAGACACTGGCTGCAGAAAGATGATATTAGTGAAGCTTACCTTACAAATCTTT 720  
QY 721 GCACCTTATCACAATGCTGCAAAATGTGCTAACTCCCTAGTCCACAGATGGCTGTACA 780  
|||||  
Db 721 GCACCTTATCACAATGCTGCAAAATGTGCTAACTCCCTAGTCCACAGATGGCTGTACA 780  
QY 781 CTGCTTCTGCTTCCCATCTGCTGACATTTGTGCAACCAAGAAATTGAATGTGGGT 840  
|||||  
Db 781 CTGCTTCTGCTTCCCATCTGCTGACATTTGTGCAACCAAGAAATTGAATGTGGGT 840  
QY 841 ATTATTGTGTGCTGAGAGACACCATCCAGGGCTTTTCACTTTCAGGACATGTGTTAC 900  
|||||  
Db 841 ATTATTGTGTGCTGAGAGACACCATCCAGGGCTTTTCACTTTCAGGACATGTGTTAC 900  
QY 901 TAACTGGGCTACTTCTCCAAAGGTTTGAACCAATTTGTTTATTTACTTATTTTGTGT 960  
|||||  
Db 901 TAACTGGGCTACTTCTCCAAAGGTTTGAACCAATTTGTTTATTTACTTATTTTGTGT 960  
QY 961 GCATGAGGTAGCATGTATACGTATGTATAGAGTCAATGATGGCTCTACCCCTCAA 1020  
|||||  
Db 961 GCATGAGGTAGCATGTATACGTATGTATAGAGTCAATGATGGCTCTACCCCTCAA 1020  
QY 1021 ATCATTTGCAGATCCCCAGCAAGTGAAGTACCGGAGCTTTGAAGTTTGTATGTGGACTG 1080  
|||||  
Db 1021 ATCATTTGCAGATCCCCAGCAAGTGAAGTACCGGAGCTTTGAAGTTTGTATGTGGACTG 1080  
QY 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGCAGCAGGGCTTAAACATGGAGCCAGCTCT 1140  
|||||  
Db 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGCAGCAGGGCTTAAACATGGAGCCAGCTCT 1140  
QY 1141 CTAGGCCCTAAGGTAATCTTTTACTTTTAAAAATATATATTTCTCA6CCGGGTGTGGTGC 1200  
|||||  
Db 1141 CTAGGCCCTAAGGTAATCTTTTACTTTTAAAAATATATATTTCTCA6CCGGGTGTGGTGC 1200  
QY 1201 ACAGCCCTTAAATCCCAGCAGCTTGAAGGCTGAGGTGTAGGAATTAATACACACAGCCAG 1260  
|||||  
Db 1201 ACAGCCCTTAAATCCCAGCAGCTTGAAGGCTGAGGTGTAGGAATTAATACACACAGCCAG 1260  
QY 1261 CTGGGGCTCAGAGCTTGGCCCTGTTTCTTTTCTTTTCTTTATGTGCACTGTGTCTTA 1320  
|||||  
Db 1261 CTGGGGCTCAGAGCTTGGCCCTGTTTCTTTTCTTTTCTTTATGTGCACTGTGTCTTA 1320  
QY 1321 CCTGGGTGTATGCTGCGTGAAGGTTGTCAGATCCCTTGGAGCTGGAGTTAAAGACAGTTG 1380  
|||||  
Db 1321 CCTGGGTGTATGCTGCGTGAAGGTTGTCAGATCCCTTGGAGCTGGAGTTAAAGACAGTTG 1380  
QY 1381 TGATACAGCTGCGCTTACAGATGCTGGAATTTGAACCCAGGTGCTCTTGAAGACAGCC 1440  
|||||  
Db 1381 TGATACAGCTGCGCTTACAGATGCTGGAATTTGAACCCAGGTGCTCTTGAAGACAGCC 1440  
QY 1441 AGTGTCTTAACTTTCAGAGCCACCCCTCAACCTGTCTTTAGAGACTTTAACTTTTG 1500  
|||||  
Db 1441 AGTGTCTTAACTTTCAGAGCCACCCCTCAACCTGTCTTTAGAGACTTTAACTTTTG 1500  
QY 1501 TGTAAATGTGGAACTGAGTGTGACATTACCAAGTGTGTGCTGCGCTGTAGACATCA 1560  
|||||

Db 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTACCAAGTGTGTGCTGCGCTGTAGACATCA 1560  
|||||  
QY 1561 CTGAGCCCGTACCACACAGCACTAATGTGATACAGTTTAAAGGCAACACTTAAATGACA 1620  
|||||  
Db 1561 CTGAGCCCGTACCACACAGCACTAATGTGATACAGTTTAAAGGCAACACTTAAATGACA 1620  
QY 1621 ATAGTTGATAGAGTTTGAATATATGCTGAGCTATTTGTTAGCCGAGCTTGTGCTGC 1680  
|||||  
Db 1621 ATAGTTGATAGAGTTTGAATATATGCTGAGCTATTTGTTAGCCGAGCTTGTGCTGC 1680  
QY 1681 TTAGCATGTGCTGTGAGAAATAGAAAAATGAAAGTCTAGTCTCTGGAACCCACA 1740  
|||||  
Db 1681 TTAGCATGTGCTGTGAGAAATAGAAAAATGAAAGTCTAGTCTCTGGAACCCACA 1740  
QY 1741 GAGGCAAGGAGAAACCCACTCTGAAAGTTGTTCTGTGAGCTTACATCAACTTACAT 1800  
|||||  
Db 1741 GAGGCAAGGAGAAACCCACTCTGAAAGTTGTTCTGTGAGCTTACATCAACTTACAT 1800  
QY 1801 AATAGTTACATGAT 1860  
|||||  
Db 1801 AATAGTTACATGAT 1860  
QY 1861 GAGATGGCTCAGCTTCCAGAGACACTTGTGCTCTTTCAGAGAGACCTAGATTCACTGCC 1920  
|||||  
Db 1861 GAGATGGCTCAGCTTCCAGAGACACTTGTGCTCTTTCAGAGAGACCTAGATTCACTGCC 1920  
QY 1921 AGGACTCATATGTGTGCTCAGAGCCATCTGTAAATTCAGTTCCAGAGGTTCCACACCT 1980  
|||||  
Db 1921 AGGACTCATATGTGTGCTCAGAGCCATCTGTAAATTCAGTTCCAGAGGTTCCACACCT 1980  
QY 1981 CTTTGGGCTCCACAGGACCAACATATAGTACAGACATTAACATGACGGCAAAACACC 2040  
|||||  
Db 1981 CTTTGGGCTCCACAGGACCAACATATAGTACAGACATTAACATGACGGCAAAACACC 2040  
QY 2041 CATACACACATTAATTAATTAAGAAACTTAAAAAGTGCATGTGTGTTGAACATTTGTCT 2100  
|||||  
Db 2041 CATACACACATTAATTAATTAAGAAACTTAAAAAGTGCATGTGTGTTGAACATTTGTCT 2100  
QY 2101 TACACATGCTGATTTGAAGCATGTACAAAGCAGACCTAAAGAGGATGTGGGGCTGAG 2160  
|||||  
Db 2101 TACACATGCTGATTTGAAGCATGTACAAAGCAGACCTAAAGAGGATGTGGGGCTGAG 2160  
QY 2161 AGATGGCTCAGGGTTAAAGAGCACTGCTCTCTTCCGAAGAGAGTCTGAGTTCAAT 2220  
|||||  
Db 2161 AGATGGCTCAGGGTTAAAGAGCACTGCTCTCTTCCGAAGAGAGTCTGAGTTCAAT 2220  
QY 2221 CCTAGCAACCAATGCTGCTCACAACATTCATTAATGAGATCTGACACCTCTCTGTGT 2280  
|||||  
Db 2221 CCTAGCAACCAATGCTGCTCACAACATTCATTAATGAGATCTGACACCTCTCTGTGT 2280  
QY 2281 GCATCTGAAGAGACGTCGAGAGCTACAGTGTACTTGAATATTAATTAATTAATTTT 2340  
|||||  
Db 2281 GCATCTGAAGAGACGTCGAGAGCTACAGTGTACTTGAATATTAATTAATTAATTTT 2340  
QY 2341 TTTAAAAAATGAAGGGATCTGAGACACCTCAAAAAGATTAAGCAGTACTCAG 2400  
|||||  
Db 2341 TTTAAAAAATGAAGGGATCTGAGACACCTCAAAAAGATTAAGCAGTACTCAG 2400  
QY 2401 GGTGATTAATCTATCTGAGATTTTCTTTCCGCTTGGCTTGCACATGGGTGACAGAGC 2460  
|||||  
Db 2401 GGTGATTAATCTATCTGAGATTTTCTTTCCGCTTGGCTTGCACATGGGTGACAGAGC 2460  
QY 2461 CCCCTTTTATTCACAAAGAACGGGTCTCATTTATTTTGAACAAAGACCTGCACT 2520  
|||||  
Db 2461 CCCCTTTTATTCACAAAGAACGGGTCTCATTTATTTTGAACAAAGACCTGCACT 2520  
QY 2521 ATGTTTACGTCTTGTGCTACTATGAGCAGGACGCGCGGCGACACACACACAC 2580  
|||||  
Db 2521 ATGTTTACGTCTTGTGCTACTATGAGCAGGACGCGCGGCGGCGACACACACACAC 2580  
QY 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAAGTCA 2640  
|||||



Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAAGTCA 2640  
QY 2641 AGAAGAGGCTGCCCTCAAAACGATCTTTCATCTTTCCCTCTAAAGAGACACAGATTC 2700  
Db 2641 AGAAGAGGCTGCCCTCAAAACGATCTTTCATCTTTCCCTCTAAAGAGACACAGATTC 2700  
QY 2701 AAGGTGGCAAAATCTACAGAGGGGGCAGAGGGAGGGGAAAGCAGGCCATGTTCC 2760  
Db 2701 AAGGTGGCAAAATCTACAGAGGGGGCAGAGGGAGGGGAAAGCAGGCCATGTTCC 2760  
QY 2761 AGAGACCTACAGAGAGGGCAGCAAGAGCAGATCCCCAGTCCAGGGCAGGAGGTGAGG 2820  
Db 2761 AGAGACCTACAGAGAGGGCAGCAAGAGCAGATCCCCAGTCCAGGGCAGGAGGTGAGG 2820  
QY 2821 CCGTGTGTCGAGAGAGAGGCGGAGAACAGGGTTCAAGGCAAGGTTTATGCA 2880  
Db 2821 CCGTGTGTCGAGAGAGAGGCGGAGAACAGGGTTCAAGGCAAGGTTTATGCA 2880  
QY 2881 GCTCATAAAAGTGGAGTCTGGCTCAGTCAAGAAAGAGAGAGAGGAAAGGCCCTTGT 2940  
Db 2881 GCTCATAAAAGTGGAGTCTGGCTCAGTCAAGAAAGAGAGAGAGGAAAGGCCCTTGT 2940  
QY 2941 GCCCAGTGAAGAGGCTGATGTAGTAGAGATCTGCAGGGGTGCCAGAGCCAC 3000  
Db 2941 GCCCAGTGAAGAGGCTGATGTAGTAGAGATCTGCAGGGGTGCCAGAGCCAC 3000  
QY 3001 CTGTCTGTCCCAAGGAAGCCCAAGTGTGAAGTCTGGCTTGGGTGCTGAGTTCAGCTA 3060  
Db 3001 CTGTCTGTCCCAAGGAAGCCCAAGTGTGAAGTCTGGCTTGGGTGCTGAGTTCAGCTA 3060  
QY 3061 CAAGACCCCGAGAGTCTACTCATCCCATCCCATGAGGCCCTGCCCCGCCACACCCCA 3120  
Db 3061 CAAGACCCCGAGAGTCTACTCATCCCATCCCATGAGGCCCTGCCCCGCCACACCCCA 3120  
QY 3121 CCCCCGACTCCCTGCGCACTTCTCTAGGGCTGGAGGGTGGCCAGCCCTGGTGGGGTTGC 3180  
Db 3121 CCCCCGACTCCCTGCGCACTTCTCTAGGGCTGGAGGGTGGCCAGCCCTGGTGGGGTTGC 3180  
QY 3181 CTACCTGTGAGAGAGCCCAAGGTTCTAGCCGGAAGTGCACCCATCCCTGAAGCTGCAGA 3240  
Db 3181 CTACCTGTGAGAGAGCCCAAGGTTCTAGCCGGAAGTGCACCCATCCCTGAAGCTGCAGA 3240  
QY 3241 GCCAAGGGGGGGCACAGGCGACGTGACAGGCTGTGCTGGGGCTCTAGGTTGCC 3300  
Db 3241 GCCAAGGGGGGGCACAGGCGACGTGACAGGCTGTGCTGGGGCTCTAGGTTGCC 3300  
QY 3301 AGGGACCTGGGCACTACTTCCCCACCCCCCATTCATCTCTCTGGGGCCATCTTCC 3360  
Db 3301 AGGGACCTGGGCACTACTTCCCCACCCCCCATTCATCTCTCTGGGGCCATCTTCC 3360  
QY 3361 CTATATATGTGAAGAACTCTGGGGGGGGGGGTGTGTGTGAGGCAAAAGTCTGTTCC 3420  
Db 3361 CTATATATGTGAAGAACTCTGGGGGGGGGGGTGTGTGTGAGGCAAAAGTCTGTTCC 3420  
QY 3421 GTCTCGACAGGAGCTTGGCACAACTCTTAAGATCTCCAGAGTGGTGCCTCTTTC 3480  
Db 3421 GTCTCGACAGGAGCTTGGCACAACTCTTAAGATCTCCAGAGTGGTGCCTCTTTC 3480  
QY 3481 CAGACAGGTAAAGCAATTGGGTGGGACACATGTTGACACAGAGTGGTTGAGAGGGACAG 3540  
Db 3481 CAGACAGGTAAAGCAATTGGGTGGGACACATGTTGACACAGAGTGGTTGAGAGGGACAG 3540  
QY 3541 GGTCTCTGCTTCTCTCTGCGCAGCTGTGCTTCTGTAGCACCTTGGATTAAGTTTGGGG 3600  
Db 3541 GGTCTCTGCTTCTCTCTGCGCAGCTGTGCTTCTGTAGCACCTTGGATTAAGTTTGGGG 3600  
QY 3601 TGAAGTAAGTGTCTGAAGCTCTGAAGAAAGCAAGAGCAGAGGCTCTTGGGGCT 3660  
Db 3601 TGAAGTAAGTGTCTGAAGCTCTGAAGAAAGCAAGAGCAGAGGCTCTTGGGGCT 3660  
QY 3661 TCAATGAAGAAATTACAGAGCCCTTTCCTTAAGTCACTTCCGTTTCATCTGTGTAG 3720  
Db 3661 TCAATGAAGAAATTACAGAGCCCTTTCCTTAAGTCACTTCCGTTTCATCTGTGTAG 3720

QY 3721 ATTCCCTGGACCAAGGTGGCTCTCTGGAGCTCAGATTTCTACAAATTAATAATCAGACAGT 3780  
Db 3721 ATTCCCTGGACCAAGGTGGCTCTCTGGAGCTCAGATTTCTACAAATTAATAATCAGACAGT 3780  
QY 3781 CCGTAGACTGGAGCTCCGCTGCTATTTACTACTTCTCTGCTGCTGCTCATTTCTGCT 3840  
Db 3781 CCGTAGACTGGAGCTCCGCTGCTATTTACTACTTCTCTGCTGCTGCTCATTTCTGCT 3840  
QY 3841 TCAATGTCTACATCTGAATGATGTTCTTGTGTACCAATTCCTCCCTGACACTCTGGGA 3900  
Db 3841 TCAATGTCTACATCTGAATGATGTTCTTGTGTACCAATTCCTCCCTGACACTCTGGGA 3900  
QY 3901 GGTGTATCTCTTGGCACATGTATCTGCGATGTAGCTGCAGCCACCAGAGAGAGGGG 3960  
Db 3901 GGTGTATCTCTTGGCACATGTATCTGCGATGTAGCTGCAGCCACCAGAGAGAGGGG 3960  
QY 3961 AGAGTACGAGAGCTGTGCTCTAGAGCCCTATTAAGCCCTGAGATCAACCCCTTCTAGAAT 4020  
Db 3961 AGAGTACGAGAGCTGTGCTCTAGAGCCCTATTAAGCCCTGAGATCAACCCCTTCTAGAAT 4020  
QY 4021 GGCCCTCCATTTTTCGGTTACCATGATCTATTTATATAGAGTGGGACGTGAAGCCA 4080  
Db 4021 GGCCCTCCATTTTTCGGTTACCATGATCTATTTATATAGAGTGGGACGTGAAGCCA 4080  
QY 4081 AACCTGCCAAGAGTTTGGACTCTACTCAGACCAAGGTATCTGCTCAGAAATCCCTG 4140  
Db 4081 AACCTGCCAAGAGTTTGGACTCTACTCAGACCAAGGTATCTGCTCAGAAATCCCTG 4140  
QY 4141 TCACTTGAAGTTGGGAGAAATCTGCTGCTGGGGCTTCCAGGTTGGTTAGCAGAGAGGT 4200  
Db 4141 TCACTTGAAGTTGGGAGAAATCTGCTGCTGGGGCTTCCAGGTTGGTTAGCAGAGAGGT 4200  
QY 4201 ATCTCTTGTATAGGGCATGACCTAGTCTATGATGTGTACTACATTCCTGTCAGTTAAAG 4260  
Db 4201 ATCTCTTGTATAGGGCATGACCTAGTCTATGATGTGTACTACATTCCTGTCAGTTAAAG 4260  
QY 4261 CTGGAAGTAAACCCACAGCGCGCCAGAGATCTCTACAGTTGTACCCCAAGAACACA 4320  
Db 4261 CTGGAAGTAAACCCACAGCGCGCCAGAGATCTCTACAGTTGTACCCCAAGAACACA 4320  
QY 4321 AGACAGTAGATATGCAAGATAGTAGTGGGGGAGGAAGAACTTAAACCCCCCAAG 4380  
Db 4321 AGACAGTAGATATGCAAGATAGTAGTGGGGGAGGAAGAACTTAAACCCCCCAAG 4380  
QY 4381 GCCCAGAGTTCGTTCCCTAGTTACAAATGCCAGTATGAGTCTAGCTATAGGCTG 4440  
Db 4381 GCCCAGAGTTCGTTCCCTAGTTACAAATGCCAGTATGAGTCTAGCTATAGGCTG 4440  
QY 4441 TGAAGTGTAGCTACAAAGCATGAGTATGTTTCAATGTGTGTGTGTATATCTGAGCAC 4500  
Db 4441 TGAAGTGTAGCTACAAAGCATGAGTATGTTTCAATGTGTGTGTGTATATCTGAGCAC 4500  
QY 4501 TTGGAGAGCTGAAGCAGAGGATTTGCTATATGTTTGAAGGCCAGCTGATATAGAGCGA 4560  
Db 4501 TTGGAGAGCTGAAGCAGAGGATTTGCTATATGTTTGAAGGCCAGCTGATATAGAGCGA 4560  
QY 4561 GACTTGTCTTTAAGAAAAAATGAAGCCACAGAGTGGGACACAGCCTTTAATCCCA 4620  
Db 4561 GACTTGTCTTTAAGAAAAAATGAAGCCACAGAGTGGGACACAGCCTTTAATCCCA 4620  
QY 4621 GCACTTGGGAGGAGAAAGCAGGAGATTTGTAGTTTCAAGGCCAGCCTGCTATATAGT 4680  
Db 4621 GCACTTGGGAGGAGAAAGCAGGAGATTTGTAGTTTCAAGGCCAGCCTGCTATATAGT 4680  
QY 4681 GAGTTCAGAGACAGCCAGGCTACACAGAGAAACCCGTTTGAAGAAACAGAAAAACA 4740  
Db 4681 GAGTTCAGAGACAGCCAGGCTACACAGAGAAACCCGTTTGAAGAAACAGAAAAACA 4740  
QY 4741 AACAAAAAACAACAAAAAACAACCAACCAACCAACCAACCTCATCTCATCTCTC 4800  
Db 4741 AACAAAAAACAACAAAAAACAACCAACCAACCAACCAACCTCATCTCATCTCTC 4800



4801 TAGCTGTGTCTGTAGTGGTAGAGTTTGGGACCTTCAGACTTATATATATAATAGGCC 4860  
|||||  
4801 TAGCTGTGTGTGTAGTGGTAGAGTTTGGGACCTTCAGACTTATATATATAATAGGCC 4860  
|||||  
4861 TTTTATACACTGTGTAGAGACGAAAGGTTTCACTGTGGACACAGTGGGACCTGAGA 4920  
|||||  
4861 TTTTATACACTGTGTAGAGACGAAAGGTTTCACTGTGGACACAGTGGGACCTGAGA 4920  
|||||  
4921 AAGTACTCTTCCAGCCAAAATTTGTGGAAAGGCTTCGTGAGAGAAATGTGTCCGAT 4980  
|||||  
4921 AAGTACTCTTCCAGCCAAAATTTGTGGAAAGGCTTCGTGAGAGAAATGTGTCCGAT 4980  
|||||  
4981 CAGACTACTGTTCTAGAAAGCGAAGAGAGGTTTGAAGAATGTTGTGGACAGACAGTT 5040  
|||||  
4981 CAGACTACTGTTCTAGAAAGCGAAGAGAGGTTTGAAGAATGTTGTGGACAGACAGTT 5040  
|||||  
5041 GGAACAGAAAGCAGAGAGGGGAGGATCCAAAGATTTCTGAACATGTAGCTTTTGGT 5100  
|||||  
5041 GGAACAGAAAGCAGAGAGGGGAGGATCCAAAGATTTCTGAACATGTAGCTTTTGGT 5100  
|||||  
5101 TCTGTGGGTGACAAAGTTCCTCCAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160  
|||||  
5101 TCTGTGGGTGACAAAGTTCCTCCAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160  
|||||  
5161 AATGATTCAGTTAGAGGACACATCCAGCCAGGGTCTTGTGCGCAAGCTAAAGAAATG 5220  
|||||  
5161 AATGATTCAGTTAGAGGACACATCCAGCCAGGGTCTTGTGCGCAAGCTAAAGAAATG 5220  
|||||  
5221 AGAGCCCTTAACCTTCCCTGAAAGTTTAGGGAGACAGAGAGCTGAGAGATCTTCTTA 5280  
|||||  
5221 AGAGCCCTTAACCTTCCCTGAAAGTTTAGGGAGACAGAGAGCTGAGAGATCTTCTTA 5280  
|||||  
5281 GGGTAAAGAGAGATATCTGCTGACCAACATGGCTAGAGACAGAGAGATGGACAG 5340  
|||||  
5281 GGGTAAAGAGAGATATCTGCTGACCAACATGGCTAGAGACAGAGAGATGGACAG 5340  
|||||  
5341 TTACCCCTCAGAACAGCCATCCCTCTGTGCTTAAGAGAGCTGGGCCCTTTCTGTTT 5400  
|||||  
5341 TTACCCCTCAGAACAGCCATCCCTCTGTGCTTAAGAGAGCTGGGCCCTTTCTGTTT 5400  
|||||  
5401 AAGATCTTACTTTTCTTCAAGAGAGAGGACAGACCTTTGTCCCTCTTGTGCTAA 5460  
|||||  
5401 AAGATCTTACTTTTCTTCAAGAGAGAGGACAGACCTTTGTCCCTCTTGTGCTAA 5460  
|||||  
5401 AAGATCTTACTTTTCTTCAAGAGAGAGGACAGACCTTTGTCCCTCTTGTGCTAA 5460  
|||||  
5461 TAAACACCCCTGTGTAAACATTTATTATTACGTGATGTTGTCTCCAGAGACGTCA 5520  
|||||  
5461 TAAACACCCCTGTGTAAACATTTATTATTACGTGATGTTGTCTCCAGAGACGTCA 5520  
|||||  
5461 TAAACACCCCTGTGTAAACATTTATTATTACGTGATGTTGTCTCCAGAGACGTCA 5520  
|||||  
5521 TCTGTAGACCTCTGCTCTTAACCTACCAAGGATGGCCACATTCCTCACCCAGAGAG 5580  
|||||  
5521 TCTGTAGACCTCTGCTCTTAACCTACCAAGGATGGCCACATTCCTCACCCAGAGAG 5580  
|||||  
5581 TGCNAAAGAGAGCCTTAGAGAAAGGTAACAGTAACAAAGATGGCCAGAGATTAACAAA 5640  
|||||  
5581 TGCNAAAGAGAGCCTTAGAGAAAGGTAACAGTAACAAAGATGGCCAGAGATTAACAAA 5640  
|||||  
5581 TGCNAAAGAGAGCCTTAGAGAAAGGTAACAGTAACAAAGATGGCCAGAGATTAACAAA 5640  
|||||  
5641 ACTACTATCTTTGTAACCAAAATTTGTTTGTGTAACAGAGAGGGGTGTGAGATGAT 5700  
|||||  
5641 ACTACTATCTTTGTAACCAAAATTTGTTTGTGTAACAGAGAGGGGTGTGAGATGAT 5700  
|||||  
5701 GT 5760  
|||||  
5701 GT 5760  
|||||  
5761 CTGTGGGACCTTTTATGCTAAAGATATCTGATATTTGGCCCATGCCAAAGGGGAT 5820  
|||||  
5761 CTGTGGGACCTTTTATGCTAAAGATATCTGATATTTGGCCCATGCCAAAGGGGAT 5820  
|||||  
5821 TGGGAGAGATCAGGCTTGTGCAACACAGTAAGCTGCCAAGATGATTTGGGCTGAA 5880  
|||||  
5821 TGGGAGAGATCAGGCTTGTGCAACACAGTAAGCTGCCAAGATGATTTGGGCTGAA 5880  
|||||  
5821 TGGGAGAGATCAGGCTTGTGCAACACAGTAAGCTGCCAAGATGATTTGGGCTGAA 5880  
|||||  
5881 TCACCAAGGGGAGGCTGATCAGAGTGAAGAACATCAAGATTAAGCCACCTGTGGG 5940  
|||||

|||||  
5881 TCACCAAGGGGAGGCTGATCAGAGTGAAGACAAATCAAGATTAAGCCACCTGTGGG 5940  
|||||  
5941 GCTCAGAAAGGAGCTTTCAAGAGCTTAAGGCGCAAGCCATTTATATTAACAGACATGAC 6000  
|||||  
5941 GCTCAGAAAGGAGCTTTCAAGAGCTTAAGGCGCAAGCCATTTATATTAACAGACATGAC 6000  
|||||  
6001 TCNAAATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGGACACC 6060  
|||||  
6001 TCNAAATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGGACACC 6060  
|||||  
6001 TCNAAATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGGACACC 6060  
|||||  
6061 TGACCTTGACATTTATAGTCACTAGGCGCAAGAGACAGTACAGAGGGTACTGGGCTTA 6120  
|||||  
6061 TGACCTTGACATTTATAGTCACTAGGCGCAAGAGACAGTACAGAGGGTACTGGGCTTA 6120  
|||||  
6061 TGACCTTGACATTTATAGTCACTAGGCGCAAGAGACAGTACAGAGGGTACTGGGCTTA 6120  
|||||  
6061 TGACCTTGACATTTATAGTCACTAGGCGCAAGAGACAGTACAGAGGGTACTGGGCTTA 6120  
|||||  
6121 CTCAGCTTGACAGGACAGTGAAGATGGGTGACCTCATCTGATGAGAGGGCTGAG 6180  
|||||  
6121 CTCAGCTTGACAGGACAGTGAAGATGGGTGACCTCATCTGATGAGAGGGCTGAG 6180  
|||||  
6181 CACCAACAGGTACAAAGTGTCCCTGTCTCATGCGCAGGATTCCTGGCAGTTTCAAG 6240  
|||||  
6181 CACCAACAGGTACAAAGTGTCCCTGTCTCATGCGCAGGATTCCTGGCAGTTTCAAG 6240  
|||||  
6180 CACCAACAGGTACAAAGTGTCCCTGTCTCATGCGCAGGATTCCTGGCAGTTTCAAG 6239  
|||||  
6241 GACTAAGAGCTATCTCTGTGGAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6300  
|||||  
6241 GACTAAGAGCTATCTCTGTGGAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6300  
|||||  
6240 GACTAAGAGCTATCTCTGTGGAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6299  
|||||  
6301 AATTAATACAGAACCCCTGGGATGAGGCTCTGACACAGAGCTTTTAAAGATC 6360  
|||||  
6301 AATTAATACAGAACCCCTGGGATGAGGCTCTGACACAGAGCTTTTAAAGATC 6360  
|||||  
6300 AATTAATACAGAACCCCTGGGATGAGGCTCTGACACAGAGCTTTTAAAGATC 6359  
|||||  
6361 CCAGGTATTCATGATCAGCAGCTGGAACAAACACAGCTCAGGTTCAACAGAAAGAGCC 6420  
|||||  
6361 CCAGGTATTCATGATCAGCAGCTGGAACAAACACAGCTCAGGTTCAACAGAAAGAGCC 6420  
|||||  
6360 CCAGGTATTCATGATCAGCAGCTGGAACAAACACAGCTCAGGTTCAACAGAAAGAGCC 6419  
|||||  
6421 AAAGTAAAGGAAAGCTTGGGATGGGAGCCTTCTTCAGGCGCAGTAGATGAGGCTGTT 6480  
|||||  
6421 AAAGTAAAGGAAAGCTTGGGATGGGAGCCTTCTTCAGGCGCAGTAGATGAGGCTGTT 6480  
|||||  
6420 AAAGTAAAGGAAAGCTTGGGATGGGAGCCTTCTTCAGGCGCAGTAGATGAGGCTGTT 6479  
|||||  
6481 AGCAGTGGGAGGCTCTCTGCTGCTGATATAGCTATCCATCCATCCATCCATCCAT 6540  
|||||  
6481 AGCAGTGGGAGGCTCTCTGCTGCTGATATAGCTATCCATCCATCCATCCATCCATCCAT 6540  
|||||  
6480 AGCAGTGGGAGGCTCTCTGCTGCTGATATAGCTATCCATCCATCCATCCATCCATCCAT 6539  
|||||  
6541 ACACCCACCCATCCATTTATGCAACCCATCTTCACATCCATCCATCCATCCATCCATCCAC 6600  
|||||  
6541 ACACCCACCCATCCATTTATGCAACCCATCTTCACATCCATCCATCCATCCATCCATCCAC 6600  
|||||  
6540 ACACCCACCCATCCATTTATGCAACCCATCTTCACATCCATCCATCCATCCATCCATCCAC 6599  
|||||  
6601 CCAGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
|||||  
6601 CCAGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
|||||  
6600 CCAGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6659  
|||||  
6661 CATTTATCCAAAGAGAACTGTATGTACTAAATGTGGAGATTTAATTAATTTTAA 6720  
|||||  
6661 CATTTATCCAAAGAGAACTGTATGTACTAAATGTGGAGATTTAATTAATTTTAA 6720  
|||||  
6660 CATTTATCCAAAGAGAACTGTATGTACTAAATGTGGAGATTTAATTAATTTTAA 6719  
|||||  
6721 ACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6780  
|||||  
6721 ACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6780  
|||||  
6720 ACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6779  
|||||  
6781 TGGCAATCGGAGAAAGGTTTGGGTGTTTCTCTTCTCCACCGGTGGGTTCTGGGA 6840  
|||||  
6781 TGGCAATCGGAGAAAGGTTTGGGTGTTTCTCTTCTCCACCGGTGGGTTCTGGGA 6840  
|||||  
6780 TGGCAATCGGAGAAAGGTTTGGGTGTTTCTCTTCTCCACCGGTGGGTTCTGGGA 6839  
|||||  
6841 TTGAATCAAAATTAATCGGGCTGTGGCAAGTGTCTTTACACCGAGCCATTTTGTGAGA 6900  
|||||  
6841 TTGAATCAAAATTAATCGGGCTGTGGCAAGTGTCTTTACACCGAGCCATTTTGTGAGA 6900  
|||||  
6840 TTGAATCAAAATTAATCGGGCTGTGGCAAGTGTCTTTACACCGAGCCATTTTGTGAGA 6899  
|||||  
6901 CATCATTTATTAAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6960  
|||||  
6901 CATCATTTATTAAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6960  
|||||  
6900 CATCATTTATTAAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6959  
|||||  
6961 CCAGGATGACCTTTAACTCTGTCTTCACCTCCACCGAGTGTAGGTTTACAGGT 7020  
|||||



KW ABCA7 transporter; ABCA7 transporter; ss.  
 OS Mus musculus.  
 XX WO200109314-A1.  
 PN 08-FEB-2001.  
 PD 28-JAN-2000; 2000WO-FR00209.  
 PF 30-JUL-1999; 99FR-0009926.  
 PR (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Chlunini G;  
 PI WPI; 2001-182953/18.  
 DR Selecting agents that modulate ABCA transporters, useful e.g. for  
 XX normalizing serum cholesterol levels, comprises using transgenic  
 PT animals with an inactive ABCA gene allele -  
 XX Claim 12; Fig 22A; 113pp; French.  
 PS The specification describes a method for selecting or screening agents  
 CC that modulate ABCA transporters. The method comprises using non-human  
 CC recombinant mammals with an inactivated allele of the gene encoding the  
 CC ABCA transporter, or cells with an inactivated allele of ABCA  
 CC transporter, from any tissue of the recombinant mammal, preferably with  
 CC an allele truncated in one or both exons corresponding to the first  
 CC and/or second ATP-binding cassettes (NBD1 or 2). Compounds that stimulate  
 CC ABCA transporters may be useful for increasing (normalizing) serum  
 CC levels of high-density lipoprotein cholesterol. The present sequence  
 CC represents the genomic sequence of a murine ABCA7 (ABCA7) transporter.  
 XX  
 SQ Sequence 23024 BP; 5195 A; 6205 C; 6348 G; 5266 T; 10 other;  
 Query Match 1.7%; Score 153.2; DB 22; Length 23024;  
 Best Local Similarity 79.1%; Pred. No. 26-26; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; Mismatches 48;  
 QY 4552 ATAGAGCAGAGCTTGTCTTTAAGAAAAATGAAAGCCACAGTGTGGCAGACGCT 4611  
 DB 16839 ACACAGAGAAACCTGTCTAATAAACCMAAACAGCGCGGACAGTGTGACAGCCT 16898  
 QY 4612 TTAATCCACGACTTGGAGGAGCAGAGCAGATTTCTGATTCAGGCCAGCCTGCT 4671  
 DB 16899 TTAATCCACGACTTGGAGGAGCAGAGCAGATTTCTGATTCAGGCCAGCCTGCT 16958  
 QY 4672 CTATAGAGTGTGATTCAGAGCAGAGGCTACACAGAAACCCGTTTGAATAACCA 4731  
 DB 16959 CTACATGATGAGTTCAGAGCAGAGCAGACTACATAAAGAACCCCTGTTTAAAAACAA 17018  
 QY 4732 GAAAAACAAACAAACAAACAAACAAACCCCAACCCCAAC 4781  
 DB 17019 ACAAACAAACACCCCTAATAAATAAATAAGTACAAACAAAC 17068  
 RESULT 4  
 AAL38337  
 ID AAL38337 standard; DNA; 215980 BP.  
 AC AAL38337;  
 XX  
 DT 15-AUG-2002 (first entry)  
 DE Complementary strand of a genomic sequence encoding a mouse Ngr3.  
 XX  
 XX Cerbrioprotective; neuroprotective; cytosolic; Nogo receptor homologue;  
 KW Ngr3; Ngr3; axonal growth; central nervous system; CNS; cerebral injury;  
 KW spinal cord injury; stroke; demyelinating disease; multiple sclerosis;  
 KW monophasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;

KW multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;  
 KW Alexander's disease; Canavan's disease; metachromatic leukodystrophy;  
 KW Krabbe's disease; immune; bait protein; genetic mapping; gene therapy;  
 KW transgenic animal; unregulated cellular growth; cancer; tumour; mouse;  
 KW murine; ds.  
 XX  
 OS Mus sp.  
 XX WO200229059-A2.  
 PN 11-APR-2002.  
 PD 06-OCT-2001; 2001WO-US31488.  
 PF 06-OCT-2000; 2000US-238361P.  
 PR (UYIA ) UNIV YALE.  
 PA (BIOJ ) BIOGEN INC.  
 XX Strilmatter SM, Cate RL, Sah DWJ;  
 PI WPI; 2002-416677/44.  
 DR Novel Nogo receptor homolog polypeptide, Ngr2 or Ngr3, useful for  
 XX treating central nervous system disorder, cerebral injury, spinal cord  
 PT injury, stroke, and demyelinating diseases -  
 XX Example 16; Page 219-275; 277pp; English.  
 PS The invention relates to a Nogo receptor homologue polypeptide, Ngr2 or  
 CC Ngr3, comprising a 50 amino acid LRKCT sequence, a 284 amino acid NLRKCT  
 CC sequence, or a 420, 461 or 392 amino acid sequence, all given in the  
 CC specification. The Ngr3 protein or its binding antibody is useful for  
 CC decreasing inhibition of axonal growth of a central nervous system (CNS)  
 CC neuron, by contacting the neuron Ngr3 or its antibody, and for treating  
 CC CNS disease, disorder or injury. Ngr3 or a vector comprising Ngr3 is  
 CC useful for treating cerebral injury, spinal cord injury, stroke,  
 CC demyelinating diseases, e.g. multiple sclerosis, monophasic  
 CC demyelination, encephalomyelitis, multifocal leukoencephalopathy,  
 CC panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,  
 CC Alexander's disease, Canavan's disease, metachromatic leukodystrophy and  
 CC Krabbe's disease. Ngr3 is useful for inducing an immune response in a  
 CC mammal against Ngr3, as a bait protein in a two-hybrid or three-hybrid  
 CC assay, and as a research tool for identification, characterisation and  
 CC purification of interacting, regulatory proteins. The nucleotide  
 CC sequences of the invention are useful for screening for RFLP associated  
 CC with certain disorders, for genetic mapping, and for gene therapy. The  
 CC vector containing Ngr3 is useful for producing non-human transgenic  
 CC animals. The Ngr3 binding antibody is useful for isolating and purifying  
 CC Ngr3, for localisation and/or quantitation of Ngr3, and for diagnostic  
 CC and therapeutic purposes. The sequences of the invention, vectors and  
 CC antibodies are useful for treating or preventing unregulated cellular  
 CC growth such as cancer and tumour growth. This polynucleotide sequence  
 CC represents the  
 SQ Sequence 215980 BP; 56872 A; 50995 C; 50240 G; 54437 T; 3436 other;  
 Query Match 1.7%; Score 152.2; DB 24; Length 215980;  
 Best Local Similarity 82.2%; Pred. No. 1,1e-25;  
 Matches 175; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
 QY 4560 AGACTTGTCTTTAAGAAAAATGAAAGCCACAGTGTGGCAGACGCTTTAATCCC 4619  
 DB 119710 ATATATTTTACTACATTAATAAACAAGCGCGGAGTGTGGCAGATACCTTTATCCC 119769  
 QY 4620 AGCACTTGGAGGAGCAGAAAGCAGGACAGATTTCTGAGTTCAAGGCCAGCTGCTATPAG 4679  
 DB 119770 AGCACTCGGAGGAGCAGAGCAGGATTTCTGAGTTGAGCCAGCCTGCTCAAGAG 119829  
 QY 4680 TGAGTTCCAGGACGACGAGGCTACACAGAAACCCGTTTGAATAACCGAAACCA 4739  
 DB 119830 TGAGTTCCAGGACGACGAGGCTACACAGAAACCTGTCTCGAATAACCAAAAAA 119889

OY 4740 AACCAAAACAAACAAACCAACCA 4772  
 DB 118990 AAAAAAAAAAAAAAAAACCACACA 119922

RESULT 5  
 AAC82696 standard; cDNA; 1445 BP.  
 AAC82696:  
 AAC82696:  
 15-MAR-2001 (first entry)  
 Murine variant Zalpha32 cDNA.  
 Zalpha32; murine; secreted alpha-helical protein-32; antiinflammatory;  
 spermatogenesis; steroidogenesis; testicular differentiation;  
 regulatory control; hypothalamic-pituitary-gonadal axis;  
 reproductive system disorder; immunological system disorder; ss.  
 Mus musculus.  
 Key Location/Qualifiers  
 CDS 19..624  
 /\*tag= a  
 /product= "variant Zalpha32"

MO200071717-A1.  
 30-NOV-2000.  
 25-MAY-2000; 2000MO-US14563.  
 26-MAY-1999; 99US-0320159.  
 (ZYMO) ZYMOGENETICS INC.  
 Conklin DC, Gao Z;  
 WPI: 2001-032038/04.  
 P-PSDB; AAB45760.

Novel mammalian secreted alpha-helical protein-32 (Zalpha32). useful for  
 identifying antagonists of its activity that are used in treating  
 Zalpha32-induced inflammation -  
 Disclosure; Page 62-63; 73pp; English.

This invention describes a novel mammalian secreted alpha-helical  
 protein-32 (Zalpha32) (I). The invention also describes (II) an isolated  
 polynucleotide (II) encoding (I); (2) an antibody (III) that specifically  
 binds to (I); and (3) an educational kit (IV) for the teaching of  
 molecular biology and/or biochemistry comprising (II). The products of  
 the invention have antiinflammatory activity. (III) as an antagonist to  
 Zalpha32 is useful for treating Zalpha32-induced inflammation. Zalpha32  
 can be used to identify inhibitors of its activity and to prepare  
 antibodies that specifically binds to Zalpha32 epitopes. Antibodies or  
 polypeptides, directly or indirectly conjugated to drugs or toxins are  
 useful for in vivo diagnostic or therapeutic applications.  
 Zalpha32-cytokine fusion proteins are useful for enhancing in vivo  
 killing of target tissue. (I) and (II) are useful to identify and isolate  
 receptors involved in spermatogenesis, steroidogenesis, testicular  
 differentiation and regulatory control of the  
 hypothalamic-pituitary-gonadal axis and also for testing disorders of the  
 reproductive system and immunological systems.

Sequence 1445 BP; 418 A; 319 C; 398 G; 310 T; 0 other:

Query Match 1.7%; Score 150.6; DB 22; Length 1445;  
 Best Local Similarity 83.4%; Pred. No. 2.1e-26;  
 Matches 171; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

4570 TTATAGAAAAAATGAAGCCAGCAGGTGGCAGCAGCTTATCCAGACCTTGGG 4629

DB 988 TTTAAAAAATGTGTGCTACCGGCGCTGTGGGCGACGCTTAAATCCAGCACTGGG 1047  
 OY 4630 AGGAGAGAGAGAGATTTCTGAGTTCAGAGCCAGCGTGTATGAGTTCAG 4689  
 DB 1048 AGGAGAGAGAGAGATTTCTGAGTTCAGAGCCAGCGTGTACGAGTTCAG 1107  
 OY 4690 GACAGCCAGGCTACACAGAAACCTGTTTGAACCAACCAACCAACCA 4749  
 DB 1108 GACAGCCAGGCTATACAGAAACCTGTTTGAACCAACCAACCAACCA 1167

OY 4750 AACCAAAACAAACCAACCAAA 4774  
 DB 1168 ACAACCAAAACCAACCAAA 1192

RESULT 6  
 AAA40866  
 ID AAA40866 standard; DNA; 7208 BP.  
 AAA40866:  
 16-AUG-2000 (first entry)  
 Murine tumour necrosis factor alpha (TNFalpha) nucleotide sequence.  
 Antisense oligonucleotide; phosphorothioate; TNFalpha; cytokine; inhibitory;  
 tumour necrosis factor alpha; inflammatory bowel disease; diabetes;  
 rheumatoid arthritis; infectious disease; multiple sclerosis; hepatitis;  
 pancreatitis; atopic dermatitis; allograft rejection;  
 autoimmune disease; inflammatory disease; ds.  
 Mus sp.  
 MO200020645-A1.  
 13-APR-2000.  
 05-OCT-1999; 99MO-US23205.  
 05-OCT-1998; 98US-0166186.  
 18-MAY-1999; 99US-0313932.  
 (ISIS-) ISIS PHARM INC.  
 Baker BF, Bennett CF, Butler MW, Shanahan WJ;  
 WPI: 2000-303808/26.  
 P-PSDB; AAY88596.

Oligonucleotide for treating diseases associated with human tumour  
 necrosis factor-alpha (TNFalpha) such as, diabetes and rheumatoid  
 arthritis, comprises nucleotide sequence complementary to intron of  
 nucleic acid encoding TNFalpha -  
 Example 8; Page 173-179; 283pp; English.

This sequence represents the murine tumour necrosis factor alpha  
 (TNFalpha) nucleotide sequence. TNFalpha is an important cytokine that  
 plays a role in host defence. It is produced mainly in macrophages and  
 monocytes in response to infection, invasion, injury or inflammation.  
 Overexpression of TNFalpha can result in disease states, particularly in  
 infectious, inflammatory and autoimmune diseases. The invention relates  
 to antisense oligonucleotides which are capable of modulating the  
 TNFalpha gene expression. The oligonucleotides optionally have a  
 phosphorothioate backbone, and may also optionally contain at least one  
 2'-O-methoxyethyl modification. The oligonucleotides are useful for  
 modulating the expression of human TNFalpha in cells and tissues,  
 reducing a human cell inflammatory response, reducing the blood glucose  
 level in a human and treating a human having a disease or condition  
 associated with TNFalpha. Examples of diseases associated with TNFalpha  
 include diabetes, inflammatory bowel disease, multiple sclerosis,  
 pancreatitis, rheumatoid arthritis, infectious disease, hepatitis, atopic

CC dermatitis or allograft rejection. The antisense oligonucleotides are  
CC also useful for modulating the function of a selected nucleic acid  
CC sequence in adipose tissue.

XX Sequence 7208 BP; 1750 A; 2011 C; 1810 G; 1637 T; 0 other;

Query Match 1.7%; Score 148.8; DB 21; Length 7208;  
Best Local Similarity 88.0%; Pred. No. 1.3e-25;  
Matches 162; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4574 AGAAAAATGAAACCCGACGAGTGGACACACGCTTTAATCCACAGCTTGGAGGC 4633  
DB 429 AAAAAAAAAAACTGGGAGTGGTGGACACACCTTTAATCCACAGCTTGGAGGC 488  
QY 4634 AGACAGGACGATTTCTGATTCAGAGTGGTGGTCTATAGAGTGGTCCAGGACA 4693  
DB 489 AGAGGACGAGGATTTCTGATTCAGAGTGGTGGTCTATAGAGTGGTCCAGGACA 548  
QY 4694 GCCAGGCTACACAGAAACCCCTGTTTGAACCAACAAACAAACAAACAAAC 4753  
DB 549 GCCAGGCTACACAGAAACCCCTGTTCTGAAAAAAGCAAAAAAAAAAAAAAAAAA 608  
QY 4754 AAAA 4757  
DB 609 AAAA 612

RESULT 7

AAF63436/c  
ID AAF63436 standard; DNA; 6789 BP.

XX AAF63436;  
XX 14-MAY-2001 (first entry)

DE Murine CD39-L4 genomic DNA sequence.

XX Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;  
KM myocardial infarction; cerebral ischemia; angina; arterial thrombosis;  
KM cerebral artery thrombosis; platelet aggregation; inflammation;  
KM apoptosis; autoimmune disorder; neurological disorder;  
KM Alzheimer's disease; Parkinson's disease; cancer; CD39-L2; ds.

OS Mus sp.

PN WO200110205-A1.

XX 15-FEB-2001.

PF 09-AUG-2000; 2000WO-US21790.

XX 09-AUG-1999; 99US-0370265.

PR 11-JAN-2000; 2000US-0481238.

PR 25-APR-2000; 2000US-0557800.

PR 26-MAY-2000; 2000US-0583231.

PR 30-JUN-2000; 2000US-0608285.

XX (HYSE-) HYSEQ INC.

XX Ford J, Mulero JJ, Yeung G;

XX WPI; 2001-147489/15.

XX Polynucleotides encoding human CD39-like polypeptides, with apyrase  
PT and/or NDPase activity, which are useful in the treatment of  
PT pathological conditions caused by thrombosis (e.g. myocardial  
PT infarction) and inflammatory disorders -

XX Example 21; Page 108-110; 203pp; English.

CC This invention relates to polynucleotides encoding human CD39-like  
CC polypeptides with apyrase and/or NDPase activity. The polypeptides having  
CC Apyrase, including NDPase, activity are useful for inhibiting platelet

CC function and can therefore be used in the prophylaxis or treatment of  
CC pathological conditions caused by or involving thrombosis or excessive  
CC coagulation or excessive platelet aggregation, such as myocardial  
CC infarction, cerebral ischemia, angina, arterial thrombosis, cerebral  
CC artery thrombosis or intracardiac thrombosis, and conditions associated  
CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
CC modulating disease states (including platelet aggregation, inflammation  
CC and apoptosis) associated with ADP or other purinergic signalling by  
CC reducing the levels of NDPs. The polypeptides are also useful for  
CC prophylaxis or treatment of inflammation related disorders, such as  
CC disorders involving sepsis or systemic inflammatory response syndrome or  
CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,  
CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;  
CC neurological disorders including neurodegenerative diseases, epilepsy,  
CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
CC sequence represents the murine CD39-L4 genomic DNA  
CC sequence.

XX Sequence 6789 BP; 1775 A; 1476 C; 1582 G; 1952 T; 4 other;

Query Match 1.6%; Score 145.6; DB 22; Length 6789;  
Best Local Similarity 62.3%; Pred. No. 8e-25;  
Matches 304; Conservative 0; Mismatches 164; Indels 20; Gaps 4;

QY 1854 GGAGGAGAGATGGCTCAGCTCCAGAGACACTGCTGCTTGGCAGAGACCTAATTC 1913  
DB 1483 GGCCACAGAGATGACGACGAAAGTATGATGCTGCTTACAGAGACCTGAGTTG 1424  
QY 1914 AGTCCACAGAG--CTCATATGGTGGCTCAGAGCATCTGTAATTCAGAGGTT 1971  
DB 1423 AGTCCACAGATGCTAGAGAGGATGTTTACTGCTTAATTCAGATCCAGAGATC 1364  
QY 1972 CCACACCTCTCTGGCTCCACAGGACCCACATATAGTACACAGACATATACGAG 2031  
DB 1363 TGACACCCCTCATGACAGTACTCACCCACACACCAATTAATAATCTCAATTG 1304  
QY 2032 CAANAACCCATACACACATTAATAATTAAGAACTTAAGAGTCACTGTTGTATA 2091  
DB 1303 AAGAAACCTAAGATAGTACATTCATTAATAATTAAGAACTTAAGAGTCACTGTT 1244  
QY 2092 CATTTGCTTACATGCTGATTTGAAGACATGTACACGACACTGAAGAGATCTG 2151  
DB 1243 AACAATGTTTTTTTTTTTTTCTAGATAGACACTAATAAATAAATAAAGT-----G 1190  
QY 2152 GGCTGGAGAGATGGCTCAGCGGTTAAGAGACTGCTCTCCGAAGAGAGTCTG 2211  
DB 1189 GGCTGTCAAGATGGCTCAGAGGTTAAGAGCACTGCTCTTTC-----GAAGGTCCTG 1134  
QY 2212 AGTTCAATCTCAGACCAACCAATGTTGCTCCACCAACCAATCAATGATGATGACCC 2271  
DB 1133 AGTTCAATCTCAGACCAACCAATGTTGCTCCACCAACCAATCAATGATGATGACCC 1074  
QY 2272 TCTTCTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2331  
DB 1073 TCTTCTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1022

QY 2332 AATCTTT 2339

DB 1021 AATCTTT 1014

RESULT 8  
AB226080  
ID AB226080 standard; DNA; 40116 BP.

XX AB226080;

XX 24-MAR-2003 (first entry)

DE Mouse DnaseX encoding genomic DNA SEQ ID NO 7.



OY		4827	GGGAGGCAGAAGCAGCGAATTTCTTGAGTTCAMAGCCACGCCCTGTCTATGTAGTAGTTC	4686
Dd		17016	GGGGGGGGAGAGAGCGGCGGATTCTTGAGTTCAGAGCCAGCGCTGTACTCAAAAGTAGATTTC	16957
OY		4687	CAGGACAGCCAGGCGCTCACAGAGAACCCCTGTTTTGAATAAACGAAAAACAACAAA	4746
Dd		16566	CTGGACAGCGCCAGGCGCTCACCAAAGAACCCGGTCTTGAATAAACAAACAACAAACAA	16897
OY		4747	ACAAACAAAACAAA 4762	
Dd		16896	ACAAACCAACCAAAA 16881	
 RESULT 10 AAAA47451/C ID AAA47451 standard; cDNA: 2221 BP. XX AC AAA47451: XX DT 20-OCT-2000 (first entry) DE Mouse TANGO 232 coding sequence.				
KM	TANGO: 221:	222:	176:	232: adipocyte; fat metabolism; obesity;
KM	cachexia; cancer;	jaundice; hepatitis; cirrhosis; cathepsin A;		
KM	hepatic disorder; neutrophil; monocyte;	renal disorder;		
KM	intestinal disorder; pituitary disorder; adrenal cortex;			
KW	hypoadrenialism; hyperadrenialism; thyroid; bone; cartilage;			
KV	osteoarthritis; mouse; ds.			
XX				
OS	Mus musculus.			
FH				
FT	Key	Location/Qualifiers		
FT	CDS	79..798		
FT		/tag= a		
FT		/product= TANGO 232		
XX				
PN	WO200039150-A2.			
XX				
PD	06-JUL-2000.			
XX				
PF	29-DEC-1999;	99WO-US31158.		
PR	30-DEC-1998;	98US-0223094.		
PA	(MILL-) MILENNIUM PHARM INC.			
PI	Sharp JD;			
DR	WPI: 2000-465732/40.			
DR	P-PSDB: AAB01418.			
PT				
PT	Novel nucleic acid molecule encoding secreted or membrane-associated proteins useful for identifying modulators and for treating disorders associated with spleen, bone, kidney, liver, pituitary and thyroid gland			
XX				
PS	Claim 1; Fig 15a-b; 129pp; English.			
XX				
CC				
CC	Human TANGO 221 and 222 nucleic acids, proteins and their modulators are useful for modulating adipocyte function and adipocyte treatment such as obesity and for treating disorders associated with abnormal fat metabolism e.g cachexia and proliferation disorders such as cancer.			
CC	Further TANGO 222 nucleic acids and proteins are useful for treating disorders associated with spleen and hepatic disorders such as jaundice, hepatitis, cirrhosis or malignant tumors. TANGO 176 nucleic acids, polypeptides and their modulators are useful for treating lysosomal protective protein cathepsin A-associated disorders such as galactosialidosis and disorders associated with a defect in neutrophil or monocytic chemotaxis. They are also useful for treating renal disorders, intestinal disorders, pituitary related disorders, adrenal cortex disorders such as hypoadrenialism, hyperadrenialism or neoplasia			

CC	and disorders of the thyroid gland. TANGO 232 nucleic acids, proteins
CC	and their modulators are useful for treating cartilage and bone
CC	associated disease and disorders such as bone cancer and
CC	osteoarthritis.
So	Sequence 2221 BP; 550 A; 549 C; 559 G; 563 T; 0 other:
Query Match	1.6%; Score 144.6; DB 21; Length 2221;
Best Local Similarity	82.9%; Pred. No. 8e-25;
Matches 165;	Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Oy	4571 TTAGGAAAAAATGAAAGCCAGCAGAGTGTTGGCAGCGCTTTATCCAGACTTGGGA 4630
Db	2148 TTGAAATAATTAACCACTGCTGGGGGGGGTGGCCATGCTTAATCCAGACTTGGGA 2089
Oy	4631 GGCAGAGAGCAGCAGATTTTCTGAGTTCAGGCCAGCCTGCTATAGAGTGAGTTCCAGG 4690
Db	2088 GGCAGAGAGCAGCGCGGATTTCTGAGTTCAGGCCAGCCTGCTCTCAAAAGTAGTTCCAGG 2029
Oy	4691 ACAGCCAGGCGCTACACAGAGAAACCCGTTTGGAAAAACGAAAAACAAACAAACAA 4750
Db	2028 ACAGCCAGAGCTACACAGAGAAACCCGTTTGGAAAAACAAACAAACAAACCCCA 1969
Oy	4751 AACCAAAACAAACCCCAAC 4769
Db	1968 AACCAAAACAGACCACTAAC 1950
RESULT 11	
AAD44193/C	
ID	AAD44193 standard; cDNA; 2221 BP.
XX	
AC	AAD44193;
XX	
DT	13-DEC-2002 (first entry)
XX	
DE	Mouse TANGO 232 cDNA.
XX	
KW	TANGO protein; vaccine; chromosomal mapping; antisense gene therapy;
KW	forensic biology; predictive medicine; cytostatic; pharmacogenomic;
XX	tissue typing; mouse; gene; ss.
OS	Mus musculus.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	79..798
FT	/*tag= a
FT	/product= "Mouse TANGO 232 protein"
FT	79..135
FT	/*tag= b
FT	136..795
FT	mat_peptide
FT	/*tag= c
FT	/product= "Mouse mature TANGO 232 protein"
XX	
XX	US2002055139-A1.
XX	
PD	09-MAY-2002.
XX	
PF	01-MAR-2001; 2001US-0796858.
XX	
PR	30-DEC-1998; 98US-0223546.
PR	23-DEC-1999; 99US-0471179.
PR	30-DEC-1998; 98US-0223094.
PR	30-DEC-1998; 98US-0223546.
PR	30-DEC-1998; 98US-0224246.
PR	14-MAY-1999; 99US-0312359.
PR	18-JUN-1999; 99US-0336536.
PR	29-JUN-1999; 99US-0342687.
PR	30-JUL-1999; 99US-0365164.
PR	20-SEP-1999; 99US-0399723.
PR	29-DEC-1999; 99US-0474071.
PR	29-DEC-1999; 99US-0474072.
PR	15-MAY-2000; 2000US-0572002.

PR 19-JUN-2000; 2000US-0597993.  
 PR 22-JUN-2000; 2000US-0595956.  
 PR 29-JUN-2000; 2000US-0606565.  
 PR 31-JUL-2000; 2000US-0630334.  
 PR 20-SEP-2000; 2000US-0665666.  
 XX  
 PA (HOLT/) HOLTZMAN D A.  
 PA (SHAR/) SHARP J D.  
 PA (LEIB/) LEIBY K R.  
 PA (BOSS/) BOSSONE S.  
 PA (PANY/) PAN Y.  
 PA (BARN/) BARNES T M.  
 PA (FRAS/) FRASER C C.  
 PA (WRIE/) WRIGHTON N.  
 PA (MYER/) MYERS P S.  
 PA (KING/) KINGSBURY G.  
 XX  
 PI Holtzman DA, Sharp JD, Lelby KR, Bossone S, Pan Y, Barnes TM;  
 PI Fraser CC, Wrihton N, Myers PS, Kingsbury G;  
 XX  
 DR WPI: 2002-453953/48.  
 DR P-PSDB: AAE26445.  
 XX  
 PT New isolated nucleic acid encoding a TANGO polypeptide, for use as a  
 PT modulating agent for regulating cellular processes and for use in a  
 PT vaccine -  
 XX  
 PS Claim 1; Page 88-89; 206pp; English.  
 XX  
 CC The invention relates to TANGO polypeptide and its corresponding  
 CC nucleic acid sequence. TANGO protein is used to identify a compound  
 CC which binds to it that can be used for modulating its activity. It  
 CC is also used to produce an antibody. The antibody is used to detect  
 CC the presence of the polypeptide in a sample. TANGO DNA and protein  
 CC are useful as modulating agents in regulating cellular processes.  
 CC They can be used in vaccines. TANGO DNA and protein and its antibody  
 CC are used in e.g. chromosomal mapping, tissue typing, forensic biology,  
 CC predictive medicine, pharmacogenomics and treatment methods. TANGO  
 CC DNA is used in antisense gene therapy. The present sequence is mouse  
 CC TANGO 232 cDNA.  
 CC  
 XX  
 SQ Sequence 2221 BP; 550 A; 549 C; 559 G; 563 T; 0 other;  
 Query Match 1.6%; Score 144.6; DB 24; Length 2221;  
 Best Local Similarity 82.9%; Pred. No. 8e-25;  
 Matches 165; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
 QY 4571 TTAAGAAAAAATGAAGCCAGCAGTGTGACACAGCCTTTAATCCAGCACTTGGGA 4630  
 DB 2148 TTACAAATTAACACACCTGCGGGGTGGTGGCCATGCTTTAATCCAGCACTTGGGA 2089  
 QY 4631 GGCAGAAAGCAGCAGATTTCTGAGTTCAAGGCCAGCTGTCTATAGATGAGTTCCAGG 4690  
 DB 2088 GGCAGAGCAGCGGATTTCTGAGTTCAAGGCCAGCTGTCTATAGATGAGTTCCAGG 2029  
 QY 4691 ACAGCCAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAAACCAAAACAA 4750  
 DB 2028 ACAGCCAGGCTACACAGAGAAACCTGTTTGAAGAAACCAAAACCAAAACCA 1969  
 QY 4751 AACAAACAAACCAACCAAC 4769  
 DB 1968 AACAAACAAACCAACCAAC 1950  
 RESULT 12  
 AAL1384  
 ID AAL1384 standard; DNA; 17135 BP.  
 XX  
 AC AAL1384;  
 XX  
 DT 03-APR-2003 (first entry)  
 XX  
 DE pgamma6.0FlExbeta-Gal plasmid DNA SEQ ID No 55.

XX  
 KW site-specific recombinase targeting sequence; SSRs; L1; L2; transgene;  
 KW site-specific stable inversion; high efficiency DNA transfer;  
 KW site-specific recombination mediated cassette exchange; gene therapy;  
 KW ds.  
 XX  
 OS Unidentified.  
 XX  
 PN W020028353-A2.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 19-APR-2002; 2002WO-IB02493.  
 XX  
 PR 27-APR-2001; 2001US-0843150.  
 XX  
 PA (ADER-) ADEREGEM ASSOC DEV RECH EN GENETIQ.  
 XX  
 PI Chambon P, Chysellinck NB, Schuetgen F;  
 XX  
 DR WPI: 2003-167118/16.  
 XX  
 PT New DNA molecules comprising a sequence A flanked by site-specific  
 PT recombinase targeting sequences L1 and L2, useful as a transgene, and  
 PT in site-specific stable inversion of a DNA sequence or recombination  
 PT mediated cassette exchange -  
 XX  
 PS Disclosure; Page 115-120; 123pp; English.  
 XX  
 CC The invention relates to an isolated DNA molecule comprising a sequence  
 CC flanked by at least site-specific recombinase targeting sequences (SSRs)  
 CC L1, and at least a sequence flanked by at least SSRs L2. The DNA  
 CC molecule or a vector comprising the DNA molecule is useful as a  
 CC transgene, and in performing site-specific stable inversion of a DNA  
 CC sequence or site-specific recombination mediated cassette exchange. The  
 CC DNA and the vector may also be used for a variety of therapeutic and  
 CC diagnostic applications which require stable and efficient integration of  
 CC transgene sequences into genomic DNA of cells (gene therapy), for  
 CC transfection of eukaryotic cells and provide the advantage of high  
 CC efficiency DNA transfer. This polynucleotide sequence represents the DNA  
 CC of the pgamma6.0FlExbeta-Gal plasmid used in the process of the  
 CC invention.  
 CC  
 XX  
 SQ Sequence 17135 BP; 4034 A; 4434 C; 4513 G; 4154 T; 0 other;  
 Query Match 1.6%; Score 143.2; DB 25; Length 17135;  
 Best Local Similarity 87.2%; Pred. No. 5e-24;  
 Matches 157; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
 QY 4590 CCAGAGTGGTGGCAGACGCTTTAATCCAGCACTTGGGAGCAGAGCAGATTT 4649  
 DB 10123 CCGGGCCTGGTGGCAGACGCTTTAATCCAGCACTTGGGAGCAGAGCAGATTT 10182  
 QY 4650 CTGAGTTCAAGCCAGCCTGTCTATAGATGAGTTCCAGAGCCAGGCTATACAGA 4709  
 DB 10183 CTGAGTTCAAGCCAGCCTGTCTATAGATGAGTTCCAGAGCCAGGCTATACAGA 10242  
 QY 4710 GAAACCTGTTTGAAGAAACCAAGAAACCAAAACCAAAACCAAAACCA 4769  
 DB 10243 GAAACCTGATTCAAAAAACCAGAAACCAAAACCAAAACCAAAACCA 10302  
 RESULT 13  
 AAL53529  
 ID AAL53529 standard; DNA; 14707 BP.  
 XX  
 AC AAL53529;  
 XX  
 DT 30-JAN-2003 (first entry)  
 XX  
 DE Genomic DNA of mouse EHD1 gene, SEQ ID No 3.  
 XX  
 KW Cytostatic; osteopathic; recombinant Eps15 homology; EH domain; EHD1;



KW EHD2; endocytosis; IGFI signaling; suppressing adhesion; invasion;  
 KW metastasis; bone formation; osteoporosis; mouse; murine; gene; ds.  
 XX  
 OS Mus sp.  
 XX  
 PN US2002115069-A1.  
 XX  
 PD 22-AUG-2002.  
 XX  
 PF 17-MAY-1999; 99US-0312762.  
 XX  
 PR 21-FEB-1997; 97IL-0120283.  
 PR 20-FEB-1998; 98US-0026898.  
 XX  
 PA (HORO/) HOROWITZ M.  
 PA (MINT/) MINTZ L.  
 XX  
 PI Horowitz M, Mintz L;  
 DR WPI; 2002-731288/79.  
 XX  
 PT Novel recombinant Eps15 homology domain containing protein useful for  
 PT regulating endogenous EHD protein by regulating insulin-like growth  
 PT factor 1 receptor cell signaling via altered clathrin coated pit  
 PT mediated endocytosis  
 XX  
 PS Claim 6: Page 32-39; 88pp; English.  
 XX  
 CC The invention relates to a novel recombinant Eps15 homology (EH) domain  
 CC containing protein 1 or 2 (EHD1 or EHD2) with an N-terminal region  
 CC containing a nucleotide binding consensus site, a central coiled coil  
 CC structure and a C-terminal region including an EH domain, where the  
 CC polypeptide participates in endocytosis. The peptides, peptide analogues  
 CC and/or the polynucleotide sequences are useful for regulating EHD1/2  
 CC protein activity in vivo, where the EHD1/2 protein and its gene are  
 CC useful for upregulating the protein activity. The novel method is useful  
 CC for lowering the rate of IGFI signaling and therefore useful for  
 CC suppressing adhesion, invasion and metastasis where EHD is overexpressed,  
 CC and therefore endocytosis. The method is also useful for elongating IGFI  
 CC effects and therefore for increasing bone formation in osteoporosis  
 CC conditions. This polynucleotide sequence represents the genomic DNA of a  
 CC mouse EHD1 gene of the invention.  
 XX  
 SQ Sequence 14707 BP; 3397 A; 3826 C; 3894 G; 3568 T; 22 other;  
 XX  
 QY Query Match 1.6%; Score 142.6; DB 24; Length 14707;  
 Best Local Similarity 75.1%; Pred. No. 6.5e-24;  
 Matches 178; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
 OY 4576 AAAAAAATGAAGCCAGCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 4635  
 DB 12047 AAGAAATGAAGCGGGGGGGGATGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 12106  
 OY 4636 AAGCAGGCGAGATTTCGAGTTAAGGCCAGCCTGCTATAGAGTGAAGTCCAGGACAGC 4695  
 DB 12107 AAGCAGGCGGAGATTTCGAGTTGAGGCCAGCCTGCTATAGAGTGAAGTCCAGGACAGC 12166  
 OY 4696 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 4755  
 DB 12167 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 12226  
 OY 4756 AACCAAAACCAAAACCAAAACCTGCTATAGAGTGAAGTCCAGGACAGC 4812  
 DB 12227 GTTGTAGACAGTGAAGGCTCCCTGCTGCTGAGTCAAGTTCGAGTTTGGGTTT 12283

XX Mouse high growth region.  
 DE  
 XX  
 KW High growth region; high growth phenotype: Socs2; body size;  
 KW suppressor of cytokine signaling 2; ds; mouse.  
 XX  
 OS Mus sp.  
 XX  
 PN US2002155564-A1.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 26-JAN-2001; 2001US-0771208.  
 PF 29-DEC-1997; 97US-0999477.  
 PR  
 XX (REGC ) UNITV CALIFORNIA.  
 PA Medrano JF, Bradford E, Horvat S;  
 PI WPI; 2003-182637/18.  
 DR  
 XX  
 PT Novel gene that when downregulated or knocked-out, results in high  
 PT growth phenotype, useful for regulating body size in mammals e.g.  
 PT rodent, bovine and canine  
 XX  
 PS Disclosure; SEQ ID NO 18; 49pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule encoding a  
 CC gene product that, when knocked out, results in a high growth (hg)  
 CC phenotype. For example a nucleic acid disrupting the Socs2 gene is  
 CC useful for producing an animal characterised by a hg phenotype, by  
 CC inhibiting expression of Socs2 (suppressor of cytokine signaling 2)  
 CC gene. The nucleic acids of the invention are useful for regulating  
 CC body size in mammals. The nucleic acids of the invention  
 CC are useful for regulating body size in mammals. This sequence  
 CC represents the mouse high growth region.  
 CC Note: This sequence did not form part of the printed specification  
 CC but was obtained in electronic format directly from the US patent  
 CC office at  
 CC seqdata.uspto.gov/sequence.html?docID=20020155564.  
 XX  
 SQ Sequence 659158 BP; 177521 A; 147222 C; 149414 G; 184754 T; 247 other;  
 XX  
 QY Query Match 1.6%; Score 142.2; DB 25; Length 659158;  
 Best Local Similarity 85.0%; Pred. No. 5.8e-23;  
 Matches 159; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
 OY 4576 AAAAAAATGAAGCCAGCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 4635  
 DB 20987 AAGAAAGAAAGCGGGGGGGGATGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 21046  
 OY 4636 AAGCAGGCGAGATTTCGAGTTAAGGCCAGCCTGCTATAGAGTGAAGTCCAGGACAGC 4695  
 DB 21047 AAGCAGGCGGAGATTTCGAGTTGAGGCCAGCCTGCTATAGAGTGAAGTCCAGGACAGC 21106  
 OY 4696 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 4755  
 DB 21107 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 21166  
 OY 4756 AACCAAA 4762  
 DB 21167 AAAAAAA 21173

RESULT 14  
 ABX16390  
 ID ABX16390 standard; DNA; 659158 BP.  
 XX  
 AC ABX16390;  
 XX  
 DT 09-APR-2003 (first entry)

RESULT 15  
 AAF57269  
 ID AAF57269 standard; DNA; 15914 BP.  
 XX  
 AC AAF57269;  
 XX  
 DT 29-MAY-2001 (first entry)

DE Mouse RetL5 genomic DNA sequence.

XX Ret ligand 5; RetL5; autophosphorylation; tumour; renal; nephrotropic;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; mouse;  
 KM vulnerability; nontoxic; anti-HIV; neuroprotective; antibacterial; ds;  
 XX cerebroprotective; hemostatic; antiinflammatory; antiviral; neuroleptic.  
 OS Mus sp.

XX WO200116169-A2.

XX 08-MAR-2001.

XX 01-SEP-2000; 2000WO-US24111.

XX 01-SEP-1999; 99US-0152024.

XX (BIOJ ) BIOGEN INC.

XX Morley D;

XX WPI; 2001-235091/24.

XX Novel Ret ligand polypeptide useful for suppressing growth of a tumor  
 PT cell that expresses Ret and for modulating Ret signal transduction  
 PT involving a cell expressing Ret polypeptide or Ret ligand polypeptide

XX Claim 1; Page 42-53; 76pp; English.

XX The invention relates to mouse and human Ret ligand 5 (RetL5)  
 CC polypeptides. The RetL5 polypeptides can be expressed by standard  
 CC recombinant methodology. The RetL5 when bound to Ret, acts as a  
 CC dimerization or autophosphorylation activator. The polypeptides and their  
 CC antibodies are useful for stimulating growth of or limiting damage to,  
 CC Ret expressing tissue in a subject, for suppressing growth of a tumour  
 CC cell that expresses Ret, for modulating Ret signal transduction involving  
 CC a cell expressing the Ret polypeptide. The RetL5 polypeptides, fusion  
 CC proteins containing RetL5 and antibodies are useful for stimulating renal  
 CC tissue growth and/or survival, supporting renal function and minimizing  
 CC damage to renal tissue after various insults, particularly for treating  
 CC acute renal failure, acute nephritis, chronic renal failure, nephrotic  
 CC syndrome, renal tubule defects, kidney transplants, toxic injury, hypoxic  
 CC injury and trauma. The compounds are also useful for treating conditions  
 CC such as neural degeneration where neural growth and regeneration are  
 CC desirable, e.g., Alzheimer's disease, Parkinson's disease, Huntington's  
 CC disease, Tourette's syndrome, amyotrophic lateral sclerosis, as well as  
 CC motor neuron disease, demyelinating disease, bacterial diseases, viral  
 CC diseases, and prion diseases including Creutzfeldt-Jakob disease. The  
 CC compounds are also useful for treating disorders due to damage to neural  
 CC tissue caused by neoplastic impingement, trauma or cerebrovascular events  
 CC such as hemorrhage or emboli, and neural disorders such as mental  
 CC retardation, autism, fetal alcohol syndrome, Down's syndrome and cerebral  
 CC palsy. The present sequence represents the genomic DNA sequence of mouse  
 CC RetL5.  
 XX

SQ Sequence 15914 BP; 3898 A; 3883 C; 4064 G; 4064 T; 5 other:

Query Match 1.6%; Score 141.8; DB 22; Length 15914;

Best Local Similarity 83.4%; Pred. No. 1.1e-23;

Matches 161; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 4562 ACTTTGCTTTAAGAAAAATGAAAGCCAGAGTGGTGGACACGCTTTAATCCAG 4621

DB 4668 ATTTCTTTTAAAAAAGAAAGCCGGCGATGGTGACATGCTTTAATCCAG 4727

OY 4622 CACTTGGAGGACAGGACAGATTCTGAGTTCAAGGCCCTGGTCTATAGAGTG 4681

DB 4728 CACTTGGAGGACAGGACAGATTCTGAGTTCAAGGCCCTGGTCTATAGAGTG 4787

OY 4682 AGTTCCAGACAGGAGGCTTACAGAGAAACCTGTTTGAAGAAACGAAACAA 4741

DB 4788 AGTTCCAGACAGGAGGCTTACAGAGAAACCTGTTTGAAGAAACGAAACGAA 4847

OY 4742 ACAAAACAAACA 4754

DB 4848 GAAAGAAAGAAAGA 4860

Search completed: September 13, 2003, 19:46:15  
 Job time : 2045 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 19:05:48 ; Search time 31328 Seconds  
(without alignments)  
11746.109 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctgtgtcaccaagaca.....ctctagctctgcaccatg 8995

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 segs, 2045481336 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:.\*  
1: gb\_ba:.\*  
2: gb\_hlg:.\*  
3: gb\_in:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vl:.\*  
15: em\_ba:.\*  
16: em\_fun:.\*  
17: em\_hum:.\*  
18: em\_in:.\*  
19: em\_mu:.\*  
20: em\_om:.\*  
21: em\_or:.\*  
22: em\_ov:.\*  
23: em\_pat:.\*  
24: em\_ph:.\*  
25: em\_pl:.\*  
26: em\_ro:.\*  
27: em\_sts:.\*  
28: em\_un:.\*  
29: em\_vl:.\*  
30: em\_hlg\_hum:.\*  
31: em\_hlg\_inv:.\*  
32: em\_hlg\_other:.\*  
33: em\_hlg\_mus:.\*  
34: em\_hlg\_pln:.\*  
35: em\_hlg\_rod:.\*  
36: em\_hlg\_man:.\*  
37: em\_hlg\_vrt:.\*  
38: em\_sy:.\*  
39: em\_higo\_hum:.\*  
40: em\_higo\_mus:.\*  
41: em\_higo\_other:.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8995	100.0	8995	6 AX027824	AX027824 Sequence
2	8762.2	97.4	192060	2 AC098570	AC098570 Mus muscu
3	6067	67.4	170074	2 AC132885	AC132885 Mus muscu
4	3116.8	34.7	193472	2 AC115173	AC115173 Rattus no
5	3016.2	33.5	149109	2 AC101018	AC101018 Rattus no
6	564.6	6.3	779	11 BV067438	BV067438 S212P6046
7	188.6	2.1	203605	2 AC119514	AC119514 Rattus no
8	188.6	2.1	234549	2 AC128412	AC128412 Rattus no
9	188.6	2.1	240435	2 AC094169	AC094169 Rattus no
10	188.4	2.1	262322	2 AC132539	AC132539 Rattus no
11	188.2	2.1	238330	2 AC114361	AC114361 Rattus no
12	188	2.1	149425	2 AC013320	AC013320 Homo sapi
13	188	2.1	191754	9 AC021016	AC021016 Homo sapi
14	186.6	2.1	568	11 G76462	G76462 S208P543FH
15	185	2.1	195636	2 AC100085	AC100085 Mus muscu
16	184.8	2.1	1220138	2 AC115965	AC115965 Mus muscu
17	184	2.0	184591	2 AC102777	AC102777 Mus muscu
18	183.6	2.0	273800	10 AF100956	AF100956 Mus muscu
19	181.8	2.0	193534	2 AC130278	AC130278 Mus muscu
20	179.2	2.0	93419	10 AL929418	AL929418 Mouse DNA
21	179.2	2.0	218558	2 AC025964	AC025964 Mus muscu
22	179.2	2.0	240536	10 AC025910	AC025910 Mus muscu
23	179.2	2.0	241400	10 AL646093	AL646093 Mouse DNA
24	178.2	2.0	198923	2 AC109263	AC109263 Mus muscu
25	177.8	2.0	209206	2 AC117095	AC117095 Rattus no
26	177.8	2.0	246991	2 AC122956	AC122956 Rattus no
27	177.6	2.0	268276	2 AC132103	AC132103 Mus muscu
28	177.4	2.0	192129	2 AC134667	AC134667 Mus muscu
29	177.4	2.0	208161	2 AC074145	AC074145 Mus muscu
30	177.2	2.0	166167	10 AL611936	AL611936 Mouse DNA
31	177.2	2.0	248677	2 AL844206	AL844206 Mus muscu
32	177	2.0	210372	10 AC122807	AC122807 Mus muscu
33	177	2.0	224766	2 AC132138	AC132138 Mus muscu
34	176.6	2.0	173425	2 AC131034	AC131034 Mus muscu
35	176.6	2.0	204316	10 AC087541	AC087541 Mus muscu
36	176.2	2.0	180526	2 AC136083	AC136083 Rattus no
37	175.8	2.0	200057	2 AC073810	AC073810 Mus muscu
38	175.6	2.0	227423	2 AC134604	AC134604 Mus muscu
39	175.6	2.0	251490	2 AC133083	AC133083 Mus muscu
40	174.4	1.9	207814	10 AL590994	AL590994 Mouse DNA
41	174.2	1.9	141166	10 AC122423	AC122423 Mus muscu
42	174	1.9	170142	2 AC140286	AC140286 Mus muscu
43	174	1.9	173567	2 AC120412	AC120412 Mus muscu
44	174	1.9	187376	2 AC138613	AC138613 Mus muscu
45	174	1.9	231350	2 AC123357	AC123357 Rattus no

## ALIGNMENTS

RESULT 1  
AX027824  
LOCUS AX027824 8995 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent WO0034492.  
ACCESSION AX027824  
VERSION AX027824.1 GI:10188668  
KEYWORDS  
SOURCE  
MUS SP.  
MUS SP.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Robine,S., Louvard,D., Pinto,D. and Jaisser,F.  
Regulatory sequences of the mouse villin gene - use in transgenesis  
Patent: WO 0034492-A 1 15-JUN-2000;  
JOURNAL

Pred. No. is the number of results predicted by chance to have a





QY	4141	TCACCTGAGTTGGAGAAATCGCTCTGGGGGCTTCCAGSTCTYTGTTAGCAGAGGGT	4200
Db	4141	TCACCTGAGTTGGAGAAATCGCTCTGGGGGCTTCCAGSTCTYTGTTAGCAGAGGGT	4200
QY	4201	ATCCTTTATAGGGCATGACCTAGTCTATGTTGTACTACATTCCTCCAGTTAAAG	4260
Db	4201	ATCCTTTATAGGGCATGACCTAGTCTATGTTGTACTACATTCCTCCAGTTAAAG	4260
QY	4261	CTGGAACTAAACCAGCGAGCGCCAGATTCTCTACAGTTGTACCCCAAGAACACA	4320
Db	4261	CTGGAACTAAACCAGCGAGCGCGCCAGATTCTCTACAGTTGTACCCCAAGAACACA	4320
QY	4321	AGACAGTATGATATGCAGAGATAGTACGTGGGAGAGAAACTTAAACCCCAAG	4380
Db	4321	AGACAGTATGATATGCAGAGATAGTACGTGGGAGAGAAACTTAAACCCCAAG	4380
QY	4381	GCCCCAGATTCCGTTCCCTAGTTCACAAATGCCAGTATGAGTGGCTACTATGGGCTG	4440
Db	4381	GCCCCAGATTCCGTTCCCTAGTTCACAAATGCCAGTATGAGTGGCTACTATGGGCTG	4440
QY	4441	TGAGTTGGTACCTACAAACATGAGTATGTTCATGTGTAGTGTATTAATCTGAGCAC	4500
Db	4441	TGAGTTGGTACCTACAAACATGAGTATGTTCATGTGTAGTGTATTAATCTGAGCAC	4500
QY	4501	TTGGGAGGCTTAACAGCAGAGAGATTGCTATATGTTTGAGGCCAGCTGAGCTATAGAGCA	4560
Db	4501	TTGGGAGGCTTAACAGCAGAGAGATTGCTATATGTTTGAGGCCAGCTGAGCTATAGAGCA	4560
QY	4561	GACCTTTGCTTTAAGAAAAAATGAAAGCCAGACAGTGGGCACACGCTTTAATCCCA	4620
Db	4561	GACCTTTGCTTTAAGAAAAAATGAAAGCCAGACAGTGGGCACACGCTTTAATCCCA	4620
QY	4621	GCACCTTGGGAGCAGAGCAGAGATTCTGAGTTCAAGCCACGCTGTCTATAGAGT	4680
Db	4621	GCACCTTGGGAGCAGAGCAGAGATTCTGAGTTCAAGCCACGCTGTCTATAGAGT	4680
QY	4681	GAGTTCCAGGACAGCAGGGCTACACAGAGAAACCCGTTTGAAGAAACCGAAGAAACAA	4740
Db	4681	GAGTTCCAGGACAGCAGGGCTACACAGAGAAACCCGTTTGAAGAAACCGAAGAAACAA	4740
QY	4741	AACAAACAAACAAACAAACCCAAACCCAAACCCCTCTCATCTCTCATCTCTC	4800
Db	4741	AACAAACAAACAAACAAACCCAAACCCAAACCCCTCTCATCTCTCATCTCTC	4800
QY	4801	TAGGCTGTCTGTCTAGTGTAGTGTAGTGTGGGACTTCACACTTATATATTAATAGCC	4860
Db	4801	TAGGCTGTCTGTCTAGTGTAGTGTAGTGTGGGACTTCACACTTATATATTAATAGCC	4860
QY	4861	TTTTTATCTAGTCAGAGCAGAGAAAGTTTCAGTCTGGGACACAGTGGGACCCCTGAGA	4920
Db	4861	TTTTTATCTAGTCAGAGCAGAGAAAGTTTCAGTCTGGGACACAGTGGGACCCCTGAGA	4920
QY	4921	AAGTACTCTTGGCAGCCCAAAATTTCTGGAAAGGCTTCTGGAAGAAAGTGTGCCGAT	4980
Db	4921	AAGTACTCTTGGCAGCCCAAAATTTCTGGAAAGGCTTCTGGAAGAAAGTGTGCCGAT	4980
QY	4981	CAGACTACTGTTCTTGAAGGCAAGAAAGAGGTTTGAAGATGTGGTGTGACACACATTT	5040
Db	4981	CAGACTACTGTTCTTGAAGGCAAGAAAGAGGTTTGAAGATGTGGTGTGACACACATTT	5040
QY	5041	GGAACAGAAAGACAGAGAGGAGGAGCATCTCAAGATTCTGAACATTTAGCTACTTTTGGT	5100
Db	5041	GGAACAGAAAGACAGAGAGGAGGAGCATCTCAAGATTCTGAACATTTAGCTACTTTTGGT	5100
QY	5101	TCTCTGGGTGACAAAGTGTCCCCAGGAGTAGGGCTGTAGAAAGGGGACAGGGGTGAGCC	5160
Db	5101	TCTCTGGGTGACAAAGTGTCCCCAGGAGTAGGGCTGTAGAAAGGGGACAGGGGTGAGCC	5160
QY	5161	AATGAGTTCAGTTGAGGAGACATCCAGCCAGGGGCTCTTGCGCAAGGTAAAGAAATG	5220
Db	5161	AATGAGTTCAGTTGAGGAGACATCCAGCCAGGGGCTCTTGCGCAAGGTAAAGAAATG	5220
QY	5221	AGACCCCTCTAACCCCTCCCTGAAAGTTTAAAGGGAACACAGAGAGCTGAGAGATCTTCTA	5280
Db	5221	AGACCCCTCTAACCCCTCCCTGAAAGTTTAAAGGGAACACAGAGAGCTGAGAGATCTTCTA	5280
QY	5281	GGGTGAAGGAGATATCTGCTTGACCAACATGGCTAGAGGACGAAGAGATTGGACAG	5340
Db	5281	GGGTGAAGGAGATATCTGCTTGACCAACATGGCTAGAGGACGAAGAGATTGGACAG	5340
QY	5341	TTACCCCTCAGAACAGACATCCCTCTTGAGGCTCTAAGAGAGCTGGGCCCCCTTCTGT	5400
Db	5341	TTACCCCTCAGAACAGACATCCCTCTTGAGGCTCTAAGAGAGCTGGGCCCCCTTCTGT	5400
QY	5401	AAGAATCTACTTTTCTTCAGAGAGAGGACAGCCCTTGTCCCCCTCCCTGTGTGCTAA	5460
Db	5401	AAGAATCTACTTTTCTTCAGAGAGAGGACAGCCCTTGTCCCCCTCCCTGTGTGCTAA	5460
QY	5461	TAAACACCCCTGTGTGTAACTTATTTTACTGTACATTTGCTCCAGAGACATGCA	5520
Db	5461	TAAACACCCCTGTGTGTAACTTATTTTACTGTACATTTGCTCCAGAGACATGCA	5520
QY	5521	TCTGTGAGACCTGTGCTTACCTAACCTACCAAGATATGGCCCATTTCTCACCCAGAGAG	5580
Db	5521	TCTGTGAGACCTGTGCTTACCTAACCTACCAAGATATGGCCCATTTCTCACCCAGAGAG	5580
QY	5581	TGCAGAGAGAGCCCTTGAGAAAGGATTAACCTAACAAGATGGCCCAAGTAAACAAAA	5640
Db	5581	TGCAGAGAGAGCCCTTGAGAAAGGATTAACCTAACAAGATGGCCCAAGTAAACAAAA	5640
QY	5641	ACTACTATCCCTTTTACCCAAATTTGTTTCTAACACAGAGAGGGGTGTGTAGTGTAT	5700
Db	5641	ACTACTATCCCTTTTACCCAAATTTGTTTCTAACACAGAGAGGGGTGTGTAGTGTAT	5700
QY	5701	GCTGT	5760
Db	5701	GCTGT	5760
QY	5761	CTTGGGGACCTTTTCAATGCTTAAAGAAATTTCTGATATTGGGCCCATGCGCAAGGGGAT	5820
Db	5761	CTTGGGGACCTTTTCAATGCTTAAAGAAATTTCTGATATTGGGCCCATGCGCAAGGGGAT	5820
QY	5821	TGGGAGAGTCAAGGCTTCTGCAACACAGTAAGCTGCCCAAGATGATGGTGGCTGAA	5880
Db	5821	TGGGAGAGTCAAGGCTTCTGCAACACAGTAAGCTGCCCAAGATGATGGTGGCTGAA	5880
QY	5881	TCACCAAGGGCAGGCTGATATCAGATGTGACAGCAATCACAAGATTAAGCCACCCGTGGG	5940
Db	5881	TCACCAAGGGCAGGCTGATATCAGATGTGACAGCAATCACAAGATTAAGCCACCCGTGGG	5940
QY	5941	GCTCAGAAAGAGAGTTTACAAAGAGTTAAAGAGCCAAAGCATTTTATTCAAAGACATGAC	6000
Db	5941	GCTCAGAAAGAGAGTTTACAAAGAGTTAAAGAGCCAAAGCATTTTATTCAAAGACATGAC	6000
QY	6001	TCAAATTCAAAGTCAAGAGAGATTTAGCTGGAGAGATGGGGCTGTACGTGTGGACACC	6060
Db	6001	TCAAATTCAAAGTCAAGAGAGATTTAGCTGGAGAGATGGGGCTGTACGTGTGGACACC	6060
QY	6061	TGACCTTGCACTTATTTAGTCACTAGGCGCAAGAGACATGACAGAGGGGAT	6120
Db	6061	TGACCTTGCACTTATTTAGTCACTAGGCGCAAGAGACATGACAGAGGGGAT	6120
QY	6121	CTCAGCTTGGAGCAGGACGTGGAAGATGGGTGACCTTCATCTGATGTGAGAGGGCTGAG	6180
Db	6121	CTCAGCTTGGAGCAGGACGTGGAAGATGGGTGACCTTCATCTGATGTGAGAGGGCTGAG	6180
QY	6181	CACCAACAGGTACAAAGTGTTCCTGTGTCTATGCGCAGAGATTCCTGGCCACTTTTCAAAG	6240
Db	6181	CACCAACAGGTACAAAGTGTTCCTGTGTCTATGCGCAGAGATTCCTGGCCACTTTTCAAAG	6240
QY	6241	GACTTAAGACTCATCTGCGTGGGAAACAAAGATTTCAAAAGCCCTAAGGCCCATTTTGGGCT	6300
Db	6241	GACTTAAGACTCATCTGCGTGGGAAACAAAGATTTCAAAAGCCCATTTTGGGCT	6300
QY	6301	AATTAAATCAGAACCCCTGGGAGTGCAGAGCTCTAGAGAGCGAGAGCTTTTAAAGAACTC	6360

|||||  
Db 6301 AATTAAATCAGAAACCCCTGGGATGTCAGGCTCTGACAGCAGAGAGCTTTTAAAGAGCTC 6360  
OY 6361 CCAGGTATCTGATCAGAGCTGGAACAAACACAGCTACAGGTTCCAAACAGAAAGAGC 6420  
Db 6361 CCAGGTATCTGATCAGAGCTGGAACAAACACAGCTACAGGTTCCAAACAGAAAGAGC 6420  
OY 6421 AAAGCTAGGAAAGCTTGGGATGAGGAGGCTTCTTCAGAGCCAGTAGATGAGAGCTGTT 6480  
Db 6421 AAAGCTAGGAAAGCTTGGGATGAGGAGGCTTCTTCAGAGCCAGTAGATGAGAGCTGTT 6480  
OY 6481 AGCAGTGTGGCAGCTTCTCTGCTGCTGTCATATACCTATTCATCCATCATTCCATCCAT 6540  
Db 6481 AGCAGTGTGGCAGCTTCTCTGCTGCTGTCATATACCTATTCATCCATCATTCCATCCAT 6540  
OY 6541 ACACCCAGCCATCCATTTATGACCCGATCCATCCATCCATCCATCCATCCATCCATCCAC 6600  
Db 6541 ACACCCAGCCATCCATTTATGACCCGATCCATCCATCCATCCATCCATCCATCCATCCAC 6600  
OY 6601 CCACGCATCCATCCAAACCT 6660  
Db 6601 CCACGCATCCATCCAAACCT 6660  
OY 6661 CATTTATCCACAGAGAACTGGTATTTACTTAATGTGGAGATTTAAATTTTATAG 6720  
Db 6661 CATTTATCCACAGAGAACTGGTATTTACTTAATGTGGAGATTTAAATTTTATAG 6720  
OY 6721 ACCTGCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
Db 6721 ACCTGCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
OY 6781 TGGCAATCGAGAAAGGTTTGGGATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840  
Db 6781 TGGCAATCGAGAAAGGTTTGGGATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840  
OY 6841 TTGAACTCAAAATTTATGCGGCTGGTGGCAAGTGTCTTACACAGCAGCAATTTTGGCAG 6900  
Db 6841 TTGAACTCAAAATTTATGCGGCTGGTGGCAAGTGTCTTACACAGCAGCAATTTTGGCAG 6900  
OY 6901 CATCATTTATTTAGAAACATCTTTATGTAGTCCAGAGCTGGCCTCAAGGTTGCTATGTG 6960  
Db 6901 CATCATTTATTTAGAAACATCTTTATGTAGTCCAGGCTGGCCTCAAGGTTGCTATGTG 6960  
OY 6961 CCACGGATGACCTTTTAATCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020  
Db 6961 CCACGGATGACCTTTTAATCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020  
OY 7021 GTTCAACTGTGTAATGCTTTAATCCACACCTGTGGGGGGGGGGGGAGGCGATC 7080  
Db 7021 GTTCAACTGTGTAATGCTTTAATCCACACCTGTGGGGGGGGGGGGAGGCGATC 7080  
OY 7081 CCTGAGTTGGAGCCAGTTTGGTCTACAGAGTTTCAGATACCTGGGGCTATACAGGAA 7140  
Db 7081 CCTGAGTTGGAGCCAGTTTGGTCTACAGAGTTTCAGATACCTGGGGCTATACAGGAA 7140  
OY 7141 ACCCTATCCCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200  
Db 7141 ACCCTATCCCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200  
OY 7201 TTAGAGGATTTAGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7260  
Db 7201 TTAGAGGATTTAGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7260  
OY 7261 ATAAATAGTAAGTACACAGATGATATCTATCTATCTATCTATCTATCTATCTATCTATCT 7320  
Db 7261 ATAAATAGTAAGTACACAGATGATATCTATCTATCTATCTATCTATCTATCTATCTATCT 7320  
OY 7321 TATCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380  
Db 7321 TATCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380  
OY 7381 CTGCTTTGACTGTAATGCTCTATATTTCTGGGTCAACCTTCAACCCCTAGTGGGTTT 7440  
|||||

|||||  
Db 7381 CTGCTTTGACTGTAATGCTCTATATTTCTGGGTCAACCTTCAACCCCTAGTGGGTTT 7440  
OY 7441 ACCACACCCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500  
Db 7441 ACCACACCCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500  
OY 7501 GGACATGAGGCTCTTGTGATGCTTAAGCAAGCTCTGCCCACAGAGCTGCAGCTCCAGTCC 7560  
Db 7501 GGACATGAGGCTCTTGTGATGCTTAAGCAAGCTCTGCCCACAGAGCTGCAGCTCCAGTCC 7560  
OY 7561 CCATTTTGTTCAGGTACTCTGTGACAGTTGTATATTTGACAGCCTATGTACTCTCTC 7620  
Db 7561 CCATTTTGTTCAGGTACTCTGTGACAGTTGTATATTTGACAGCCTATGTACTCTCTCTC 7620  
OY 7621 CACCTCCAGTTCCACACTTCTGTGTCATCCAGAGGGGGGGGCACTGTGCTCACA 7680  
Db 7621 CACCTCCAGTTCCACACTTCTGTGTCATCCAGAGGGGGGGGCACTGTGCTCACA 7680  
OY 7681 GTGCCCCGTCTCCCTGCTTCAGACCTACATATTTGCTGTGACAGTTCAATGTAATG 7740  
Db 7681 GTGCCCCGTCTCCCTGCTTCAGACCTACATATTTGCTGTGACAGTTCAATGTAATG 7740  
OY 7741 GGATGCTGCTCTGTATTTCTTTTATGCTGGCCCCCTTTATCTTTAGCAGAGTTGTGTTG 7800  
Db 7741 GGATGCTGCTCTGTATTTCTTTTATGCTGGCCCCCTTTATCTTTAGCAGAGTTGTGTTG 7800  
OY 7801 GGCCATGTGCACTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
Db 7801 GGCCATGTGCACTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
OY 7861 TGTGTGATTAACACATTTCTGTCTTCAATTTACATGATGAAATTTGGCCCCACCCAC 7920  
Db 7861 TGTGTGATTAACACATTTCTGTCTTCAATTTACATGATGAAATTTGGCCCCACCCAC 7920  
OY 7921 CCTTTTTTTTTTATTTAGACAGAGCTTTTCTGTATATCTTTGCAATCTGGCTGTCT 7980  
Db 7921 CCTTTTTTTTTTATTTAGACAGAGCTTTTCTGTATATCTTTGCAATCTGGCTGTCTCT 7980  
OY 7981 GAGCTACCTGTGACAGCAGGCTGTGAGGCTGCTCCACTTTGACACTCTGTGTGAC 8040  
Db 7981 GAGCTACCTGTGACAGCAGGCTGTGAGGCTGCTCCACTTTGACACTCTGTGTGAC 8040  
OY 8041 AGAGTAGCCATGAACCTTCAAGACAAATTTCTGTTTGTGTTTATTTACATTTGTG 8100  
Db 8041 AGAGTAGCCATGAACCTTCAAGACAAATTTCTGTTTGTGTTTATTTACATTTGTG 8100  
OY 8101 TGTATGCTGTATATGTGATGCTTTGTGTCTTATGAGTGTGACATGTGTGTGTG 8160  
Db 8101 TGTATGCTGTATATGTGATGCTTTGTGTCTTATGAGTGTGACATGTGTGTGTG 8160  
OY 8161 TGGGACAGAGAAACCAAGCATGTGCCATTCCTCAGATACAGCATCTGTTAATATGTA 8220  
Db 8161 TGGGACAGAGAAACCAAGCATGTGCCATTCCTCAGATACAGCATCTGTTAATATGTA 8220  
OY 8221 TGTATATGTATGTTTATTTAGTGTGCCAAGTATGAGATATTTTGTGAGATTTTCAC 8280  
Db 8221 TGTATATGTATGTTTATTTAGTGTGCCAAGTATGAGATATTTTGTGAGATTTTCAC 8280  
OY 8281 CTTCCTTGTGGGCTCTCCGCAATTAATCTAGCTCTGGGGCTAGTGAGCAATGCTTCA 8340  
Db 8281 CTTCCTTGTGGGCTCTCCGCAATTAATCTAGCTCTGGGGCTAGTGAGCAATGCTTCA 8340  
OY 8341 CTGATGAGCAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400  
Db 8341 CTGATGAGCAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400  
OY 8401 CGCAGTCACTGGCTTAAAGCTCACCAAGTCACTCAGATGAGTGGCTAGCCAGGAGACTCAG 8460  
Db 8401 CGCAGTCACTGGCTTAAAGCTCACCAAGTCACTCAGATGAGTGGCTAGCCAGGAGACTCAG 8460  
OY 8461 GGATATGCTGGGCTGTGCTCCACAGTGTAGAAATTTACAGGATACATCATCTGTGGAAG 8520  
Db 8461 GGATATGCTGGGCTGTGCTCCACAGTGTAGAAATTTACAGGATACATCATCTGTGGAAG 8520

QY	8521	ATTTTTAAACCGATGATCCGAGAGATAGAGACAGACACTACCAATGAGAGGGTCTTTTGT	8580
Db	8521	ATTTTAAACCGAATCTCGAGAGATAGAGACAGACACTTACCAATGAGAGGGTCTTTTGT	8580
QY	8581	GTTTGGTTGGTTCCTCTGCAATGATCAGGCAAGTCTGAATATAGTAGAGCCCTGGGCTAC	8640
Db	8581	GTTTGGTTGGTTCCTCTGCAATGATCAGGCAAGTCTGAATATAGTAGAGCCCTGGGCTAC	8640
QY	8641	ATAACATCTGTCACAAAAAGCCATATAGAGGTAGGAGAGCTGAGAGCTTAAGAAAGCCTT	8700
Db	8641	ATAACATCTGTCACAAAAAGCCATATAGAGGTAGGAGAGCTGAGAGCTTAAGAAAGCCTT	8700
QY	8701	AAGCCGGCTGTGATAGACACACAGGATAGCCCTGACACTATATAGCAAGAAGCTTTTCAAAA	8760
Db	8701	AAGCCGGCTGTGATAGACACACAGGATAGCCCTGACACTATATAGCAAGAAGCTTTTCAAAA	8760
QY	8761	ACATGAGAGGAGGGGCTATGTTTAAAGTCTGAGCTGTATACAGGACACTAAGAGCCAA	8820
Db	8761	ACATGAGAGGAGGGGCTATGTTTAAAGTCTGAGCTGTATACAGGACACTAAGAGGCCAA	8820
QY	8821	TGTAGACATTTGATAGTAAAGAGATCATCATCAAGCCGGGTGGGACAGGGTAGAGGTTGG	8880
Db	8821	TGTAGACATTTGATAGTAAAGAGATCATCATCAAGCCGGGTGGGACAGGGTAGAGGTTGG	8880
QY	8881	ACTACAGTGTCAAGACCCCCCATAGGAAGCCAGTTTCCCTTCTCTCTGAGGCTCAAGC	8940
Db	8881	ACTACAGTGTCAAGACCCCCCATAGGAAGCCAGTTTCCCTTCTCTCTGAGGCTCAAGC	8940
QY	8941	CTGGCTGAGAGGGGACAGCTCTCAAGTCCCTTCTCTAGGCTCGTGCACACATAG	8995
Db	8941	CTGGCTGAGAGGGGACAGCTCTCAAGTCCCTTCTCTAGGCTCGTGCACACATAG	8995

LOCUS	192060 bp	DNA	linear	HTG 20-AUG-2002
DEFINITION	AC0986570 Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered pieces.			
ACCESSION	AC0986570			
VERSION	AC0986570.2 GI:22325297			
KEYWORDS	HTG; HTGS; PHASE1; HTGS; DRAFT; HTGS_FULLTOP.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Euarystota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 192060)			
TITLE	Birren,B., Nusbaum,C. and Lander,E.			
JOURNAL	Mus musculus, clone RP23-278N11			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 192060)			
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Baran,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campolinos,A., Chang,Y., Chazaro,B., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Gaiagan,J., Gardyna,S., Ginde,R., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLennan,C., MacDonald,P., Major,J., Margus,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Melrim,J., Menes,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Plerre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schpack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strausman,A., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J., Tophan,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			

```

TITLE
JOURNAL
REFERENCE
AUTHORS
DIRECT Submission
Submitted (24-OCT-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192060)
Britten,B., Nushbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barns,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhallal,B.,
Camarata,J., Chang,J., Chazaro,B., Choepey,Y., Collymore,A.,
Cook,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fark,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horion,L., Hume,W., Iliev,T., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,D., Matthews,C.,
McCarthy,M., Meldrum,J., Menusil,L., Mhova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norb,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 20, 2002 this sequence version replaced g1:16356898.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14796
Center clone name: 278_N_11
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Assembly program: Dye-terminator Big Dye; 100% of reads
Consensus quality: Phrap; version 0.960731
Consensus quality: 190464 bases at least Q40
Consensus quality: 191061 bases at least Q30
Consensus quality: 191345 bases at least Q20
Insert size: 163000; agarose-1p
Insert size: 191560; sum-of-contigs
Quality coverage: 12.4 in Q20 bases; agarose-1p
Quality coverage: 10.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4921: contig of 4921 bp in length
* 4922 5021: gap of 100 bp
* 5022 15200: contig of 10179 bp in length
* 15201 15300: gap of 100 bp
* 15301 78692: contig of 63392 bp in length
* 78693 78792: gap of 100 bp
* 78793 116118: contig of 37336 bp in length
* 116119 116218: gap of 100 bp
* 116219 174491: contig of 58273 bp in length
* 174492 174591: gap of 100 bp
* 174592 192060: contig of 17469 bp in length.
*
* location/Qualifiers
* 1. 192060
*
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
FEATURES
source

```





D	41961	AATGTTACATGATGTAATTAATTAATTAGTAATTTCTTTAGAGAGTATATGTTGGAGGA	42020
O	1861	GAGATGGCTCAGCTTCCAGAGACACTTGTGCTCTTGACAGACCTAGATTCCAGTTCC	1920
D	42021	GAGATGGCTCAGCTTCCAGAGACACTTGTGCTCTTGACAGAGCTTAGATTCCAGTTCC	42080
O	1921	AGGATCTATATGATGGGCTCACAGGCATCTGTAATTCAGATTCCAGAGGTTCCACACCT	1980
D	42081	AGGATCTATATGATGGGCTCACAGGCATCTGTAATTCAGATTCCAGAGGTTCCACACCT	42140
O	1981	CTTGTGGCTCCACAGGCACCATACATAGTATACAGACATATACAGGCAAAACACC	2040
D	42141	CTTGTGGCTCCACAGGCACCATACATAGTATACAGACATATACAGGCAAAACACC	42200
O	2041	CATCACACATTAATTAATTAAGAACTTAATTAAGGTCATGTTGGTAAACATTTGCT	2100
D	42201	CATCACACATTAATTAATTAAGAACTTAATTAAGGTCATGTTGGTAAACATTTGCT	42260
O	2101	TACACATGCTGATTGAAGACATGTACACAGCACACTGAAAGAGGATCTGGGCTGAG	2160
D	42261	TACACATGCTGATTGAAGACATGTACACAGCACACTGAAAGAGGATCTGGGCTGAG	42320
O	2161	AGATGGCTCAGCGGTTAAGAGCACTGCTCTCTCCGAAGGAAGTCTTGAGTTCAAT	2220
D	42321	AGATGGCTCAGCGGTTAAGAGCACTGCTCTCTCCGAAGGAAGTCTTGAGTTCAAT	42380
O	2221	CCTAGCAACACATAGTGGCTCACACACATCCATATAGATCGACACCTCTTGCT	2280
D	42381	CCTAGCAACACATAGTGGCTCACACACATCCATATAGATCGACACCTCTTGCT	42440
O	2281	GCATCTGAAGCAGCTGCAGAGCTACAGTGTACTTATATATCTAATTAATCTTTT	2340
D	42441	GCATCTGAAGCAGCTGCAGAGCTACAGTGTACTTATATATCTAATTAATCTTTT	42500
O	2341	TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTGTGAGCATGCTACG	2400
D	42501	TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTGTGAGCATGCTACG	42560
O	2401	GGTGAATTAATCTGAGAGTTTCTCTTCCGCTTGCAACTGGGTGACAGAGC	2460
D	42561	GGTGAATTAATCTGAGAGTTTCTCTTCCGCTTGCAACTGGGTGACAGAGC	42620
O	2461	CCCTTTTCATTCACAAAGAGGGTGCATATTCTGAAACAAACACACACTGGAGT	2520
D	42621	CCCTTTTCATTCACAAAGAGGGTGCATATTCTGAAACAAACACACACTGGAGT	42680
O	2521	ATGTTTACTGTCCTGCTGACTATGAGCAGCGCAGCGCGCGCG--CACACACACAC	2578
D	42681	ATGTTTACTGTCCTGCTGACTATGAGCAGCGCAGCGCGCGCGCGCGCACACACAC	42740
O	2579	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGT	2638
D	42741	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGT	42800
O	2639	CAGAAGAGGCTGCCCTCAAAACAGATCTTATCTTCCCTCTTAAGAGACACATTT	2698
D	42801	CAGAAGAGGCTGCCCTCAAAACAGATCTTATCTTCCCTCTTAAGAGACACATTT	42860
O	2699	CCAAGGTGGCAGAAGATCTACAGGGGACAGGAGGAGGAGGAGGAGGAGGAGGATTT	2758
D	42861	CCAAGGTGGCAGAAGATCTACAGGGGACAGGAGGAGGAGGAGGAGGAGGAGGATTT	42920
O	2759	CCAGAGACCTACAGACAGAGGACAGAGGAGATCCCAAGTCTCCAGGGCAGAGGATGGA	2818
D	42921	CCAGAGACCTACAGACAGAGGACAGAGGAGATCCCAAGTCTCCAGGGCAGAGGATGGA	42980
O	2819	GGCCCTTGCTCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTTG	2878
D	42981	GGCCCTTGCTCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTTG	43040
O	2879	CAGCTCATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTT	2938
D	43041	CAGCTCATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTT	43100
O	2939	GTGCCACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTT	2998
D	43101	GTGCCACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTT	43160
O	2999	ACCTGTCTGCTCCCAAGGAAACCCCAAGTGTGAATCTGGCTTGGGTGTGAGTTCCAC	3058
D	43161	ACCTGTCTGCTCCCAAGGAAACCCCAAGTGTGAATCTGGCTTGGGTGTGAGTTCCAC	43220
O	3059	TACAGACCCACAGAGTCTTACTCATCCCATCCAGTGCCTTCGCCGCACACCC	3118
D	43221	TACAGACCCACAGAGTCTTACTCATCCCATCCAGTGCCTTCGCCGCACACCC	43280
O	3119	CACCCCGAGCTCCCGGACCATCTCTAGGGGCTGGAGGGTGGCAGCCCTGGTGGGGT	3178
D	43281	CACCCCGAGCTCCCGGACCATCTCTAGGGGCTGGAGGGTGGCAGCCCTGGTGGGGT	43340
O	3179	GCCTACCTGACAGTAGAGCCAGAGTCTTAGCGGAGTGCACCCCATCCCTGAAGCTGCA	3238
D	43341	GCCTACCTGACAGTAGAGCCAGAGTCTTAGCGGAGTGCACCCCATCCCTGAAGCTGCA	43400
O	3239	GAGCCAAAGGGGGGACACAGGAGCTCAGGCTGTACAGGCTTGTGGGCTGAGTTTC	3298
D	43401	GAGCCAAAGGGGGGACACAGGAGCTCAGGCTGTACAGGCTTGTGGGCTGAGTTTC	43460
O	3299	CCAGGAGCTGGGACACTTATCCACACCCCATCCATCTCTGCGGGCCATATTT	3358
D	43461	CCAGGAGCTGGGACACTTATCCACACCCCATCCATCTCTGCGGGCCATATTT	43520
O	3359	CCCTTATATGATGAGGAAGTCTTGAGGGGGGGGGGGGGGGGTGAGACAAAGTCTTT	3418
D	43521	CCCTTATATGATGAGGAAGTCTTGAGGGGGGGGGGGGGGGGGTGAAGAGGCTTT	43578
O	3419	CGGCTCTGACAGCAGTCTGCACAACTCTTAATATCTCCAGAGTGGTGGCTGCTCT	3478
D	43579	CGGCTCTGACAGCAGTCTGCACAACTCTTAATATCTCCAGAGTGGTGGCTGCTCT	43638
O	3479	TCCAGACAGGTAAGGCAATGGGTGGGGACATGCTGACACACAGTGGTGGAGGGAC	3538
D	43639	TCCAGACAGGTAAGGCAATGGGTGGGGACATGCTGACACACAGTGGTGGAGGGAC	43698
O	3539	AGGGCTCTGCTCTGCGGACAGCTGCGTCTCTGAGCACTTGGATTAAGTTTGGG	3598
D	43699	AGGGCTCTGCTCTGCGGACAGCTGCGTCTCTGAGCACTTGGATTAAGTTTGGG	43758
O	3599	GGTGAAGTAAAGTCTCTGAAACTCTGAAAGAAAGGAGGAGGAGGAGGAGGAGGCTT	3658
D	43759	GGTGAAGTAAAGTCTCTGAAACTCTGAAAGAAAGGAGGAGGAGGAGGAGGAGGCTT	43818
O	3659	CTTCATTAAGGAAGTTCACAGACCCCTTCTCTGAATCTCACTTCCGTCATCTGCT	3718
D	43819	CTTCATTAAGGAAGTTCACAGACCCCTTCTCTGAATCTCACTTCCGTCATCTGCT	43878
O	3719	AGATTCCTGGGACACAAAGTGGCTCTGAGGAGCTCAGATTTCTAATTAATCAGAGCA	3778
D	43879	AGATTCCTGGGACACAAAGTGGCTCTGAGGAGCTCAGATTTCTAATTAATCAGAGCA	43938

OY	4019	ATGGCCCCCATTTTTTGGTTACATGATCTATTTTTATATCAGAGTGGCGAGTAAAGC	4078
Db	44179	ATGGCCCCCATTTTTTGGTTACCGATGATCTATTTTTATATCAGAGTGGCGAGTAAAGC	44238
OY	4079	CAAAACGCGCCACAGAGTTTGGGACTACATCAGCCAAAGGTTATTCGCTCAGAAATCCCC	4138
Db	44239	CAAAACGCGCCACAGAGTTTGGGACTACATCAGCCAAAGGTTATTCGCTCAGAAATCCCC	44298
OY	4139	TGTCACCTTGAAGTTGGAGCAATTCGCTCGGGGCGCTCCAGGTCTTGGTTAGCAGAGAG	4198
Db	44299	TGTCACCTTGAAGTTGGAGCAATTCGCTCGGGGCGCTCCAGGTCTTGGTTAGCAGAGAG	44358
OY	4199	GTATCTTGTGTATAGGGCATGACCTAGTCTATGTGTGTACTCTACATTCCTGTCCAGTTAAA	4258
Db	44359	GTATCTTGTGTATAGGGCATGACCTAGTCTATGTGTGTACTCTACATTCCTGTCCAGTTAAA	44418
OY	4259	AGCTGGAACGTAAACCCACGGCGCCAGGATTCCTACATGCTTACCCCAAGAACAA	4318
Db	44419	AGCTGGAACGTAAACCCACGGCGCCAGGATTCCTACATGCTTACCCCAAGAACAA	44478
OY	4319	CAAGACAGTATGATCAAGATAGGATAGCTGGGGAGAAAGAACTTAAACCCCCCAA	4378
Db	44479	CAAGACAGTATGATCAAGATAGGATAGCTGGGGAGAAAGAACTTAAACCCCCCAA	44538
OY	4379	AGCCCCACAGTTCCTCCCTTACATTCACATGCCAGTATGAGTGTAGCTACTATGGGC	4438
Db	44539	AGCCCCACAGTTCCTCCCTTACATTCACATGCCAGTATGAGTGTAGCTACTATGGGC	44598
OY	4439	TGTAGATTGGATGATCAACATGAGAGATGTTCACTGATGATAGTGTATTAATCTGAGC	4498
Db	44599	TGTAGATTGGATGATCAACATGAGAGATGTTCACTGATGATAGTGTATTAATCTGAGC	44658
OY	4499	ACTTGGAGAGCTGAAGCAGAGAGATTCCTATATGTTTGAAGCCAGCCTAGCTATAGAGC	4558
Db	44659	ACTTGGAGAGCTGAAGCAGAGAGATTCCTATATGTTTGAAGCCAGCCTAGCTATAGAGC	44718
OY	4559	GAGACTTGTCTTAAAGAAAAAATGAAGCCAGCATGTGTGTGCGACACAGCCTTTAATCC	4618
Db	44719	GAGACTTGTCTTAAAGAAAAAATGAAGCCAGCATGTGTGTGCGACACAGCCTTTAATCC	44778
OY	4619	CAGACACTTGGAGAGCAGACAGGAGATTCCTGAGTTCAAGGCCAGCCTGCTATATGA	4678
Db	44779	CAGACACTTGGAGAGCAGACAGGAGATTCCTGAGTTCAAGGCCAGCCTGCTATATGA	44838
OY	4679	GTGAGTTCAGGACAGCGAGGGCTACACAGAGAAACCTGTTTGAAGAAACCAAGAAAAAC	4738
Db	44839	GTGAGTTCAGGACAGCGAGGGCTACACAGAGAAACCTGTTTGAAGAAACCAAGAAAAAC	44898
OY	4739	AAAAACAACAAAAACAAAAAACCACCAACCCAAACCCAAACCTCTCTCTATCTC	4798
Db	44899	AAAAACAACAAAAACAAAAAACCACCAACCCAAACCCAAACCTCTCTCTATCTC	44958
OY	4799	TCTAGCGTGTGTCTGTCTAGTGTGTAGTGTGGAGCTTCAGACTTATATATTAATATAG	4858
Db	44959	TCTAGCGTGTGTCTGTCTAGTGTGTAGTGTGGAGCTTCAGACTTATATATTAATATAG	45018
OY	4859	CCTTTATATCATCTGTCAGAGACGAGAAAGTTCAGTCTGGGACACAGTGGGACCTTGA	4918
Db	45019	CCTTTATATCATCTGTCAGAGACGAGAAAGTTCAGTCTGGGACACAGTGGGACCTTGA	45078
OY	4919	GAAAGTCTCCTTGGCCAGGCCAAATTCGCGGAAGGCTTCTCGGAGGAAGTGTCTCCG	4978
Db	45079	GAAAGTCTCCTTGGCCAGGCCAAATTCGCGGAAGGCTTCTCGGAGGAAGTGTCTCCG	45138
OY	4979	A-----	4979
Db	45139	ATTGAGTCTGGAGAAATATGAAGAGTGGGAAAGCCAAAGCAGAAAGCAAGCCTGGGTGA	45198
OY	4880	-----TCAGACTACTGTGTTCTAGAAAGCAGAAAGAGAGGGTTGGAAGATGTTGGTG	5029
Db	45199	AAGGCACAGCTCAGAGACTACTGTTCTGTGAAGGCGAGAAAGAGAGGGTTGGAAGATGTTGGTG	45258

OY	5030	TACGAGCAGTGTGGAAACAAGAAGACAGAGGGGGGAGCATCCAAATTCCTGAACATTGAGC	5089
Db	45259	GACACACAGTTTGGAACACAAAGACAGAGGGGGGAGCATCCAAATTCCTGAACATTGAGC	4531
OY	5090	TGACTTTTGGTTCCTCGGTGCACAAAGTGTCCCCAGGATATGGGCTGTAGAAAAGGAGCC	5149
Db	45319	TGACTTTTGGTTCCTCGGTGCACAAAGTGTCCCCAGGATATGGGCTGTAGAAAAGGAGCC	4537
OY	5150	AGGGGTGAGCCAATGATGTCAAGTTGAGGGACACATCCAGGCCAAGGCTCTTGCTGGCAA	5209
Db	45379	AGGGGTGAGCCAATGATGTCAAGTTGAGGGACACATCCAGGCCAAGGCTCTTGCTGGCAA	4543
OY	5210	GCTAAAGATGAGAGCCCTCTTAACCCTCCCTGAAAGTTTAGGGGAGACAGAGACTGAGG	5269
Db	45439	GCTAAAGATGAGAGCCCTCTTAACCCTCCCTGAAAGTTTAGGGGAGACAGAGACTGAGG	4549
OY	5270	AGATCCCTCTAGGGGTGAAGGAGAGGATCTGCTGACCAACATGGGCTAGGAGAGAGAGC	5329
Db	45499	AGATCCCTCTAGGGGTGAAGGAGAGGATCTGCTGACCAACATGGGCTAGGAGAGAGAGC	4555
OY	5330	AGTTGGACAGTAAACCCCTCAGAACACAGCACCCCTCTTGCTCTAAGAGAGCTGGGCC	5389
Db	45559	AGTTGGACAGTAAACCCCTCAGAACACAGCACCCCTCTTGCTCTAAGAGAGCTGGGCC	4561
OY	5380	CCTTTCCTTTAAGATCTTACTTTCTTCAGAGAGAGACGACAGCCCTTTGTCCCTCC	5449
Db	45619	CCTTTCCTTTAAGATCTTACTTTCTTCAGAGAGAGACGACAGCCCTTTGTCCCTCC	4567
OY	5450	CTGTGTGTCATAAACAACCCCTGAGTGAACATTAGTTATTTTACGTGCTGATTGGCTCC	5509
Db	45679	CTGTGTGTCATAAACAACCCCTGAGTGAACATTAGTTATTTTACGTGCTGATTGGCTCC	4573
OY	5510	AGGACAGTCACATCTGGTGTAGACCTCGCTCACTCAACCAAGATATGCCACATTCCTC	5569
Db	45739	AGGACAGTCACATCTGGTGTAGACCTCGCTCACTCAACCAAGATATGCCACATTCCTC	4579
OY	5570	AACCGAAGAGTCGAGAAAGAGACCTTATAGAAAAAGGTAAACATGTAACAAGATGGCCAGA	5629
Db	45739	AACCGAAGAGTCGAGAAAGAGACCTTATAGAAAAAGGTAAACATGTAACAAGATGGCCAGA	4585
OY	5630	ATTAACAACAAAACATCATCTCCCTTGTACCCAAATTGGTTGGCGAANCAGAGAGGGGTG	5689
Db	45859	ATTAACAACAAAACATCATCTCCCTTGTACCCAAATTGGTTGGCGAANCAGAGAGGGGTG	4591
OY	5690	TGTGAGTATAGTG	5749
Db	45919	TGTGAGTAT-----TG	4597
OY	5750	TGTGTGTGTGTGTGGGGGACTTTTCATGCTAAAGAAATATCTGATATTGGGCCCATGCC	5809
Db	45971	TGTGTGTGTGTGTGGGGGACTTTTCATGCTAAAGAAATATCTGATATTGGGCCCATGCC	4603
OY	5810	AACAGGGTATTTGGGAGAGTCAGGCTCTGC AAAACACATGTAACCTGCCCAAGATGATTT	5869
Db	46031	AACAGGGTATTTGGGAGAGTCAGGCTCTGC AAAACACATGTAACCTGCCCAAGATGATTT	4609
OY	5870	GGTGGCCCTGAATACACAAGGGGACAGGCTGATCAGAGTGGAGACAACATCAACAATTAAGC	5929
Db	46091	GGTGGCCCTGAATACACAAGGGGACAGGCTGATCAGAGTGGAGACAACATCAACAATTAAGC	4615
OY	5930	CACCTGTGGGGCTCAGAAAGGAGGATTTACAAAGAGTTAAAGGCCCAAGCCATTTATTATC	5989
Db	46151	CACCTGTGGGGCTCAGAAAGGAGGATTTACAAAGAGTTAAAGGCCCAAGCCATTTATTATC	4621
OY	5990	CAGAAGATGACTCAAAAATCAAAGTGC AAGAGAGATTATAGCTGGAGAGATGGGGCTGTACAG	6049
Db	46211	CAGAAGATGACTCAAAAATCAAAGTGC AAGAGAGATTATAGCTGGAGAGATGGGGCTGTACAG	4627
OY	6050	TGTGGGAACCTGAGACCTTGCACTTAATTAGTACAGGCAAGAGACAGTCAACAAGGGTG	6109
Db	46271	TGTGGGAACCTGAGACCTTGCACTTAATTAGTACAGGCAAGAGACAGTCAACAAGGGTG	4633
OY	6110	ACTGGGTCTACTAGCTTTGGAGCAGGCACGTGAGAGATGGGTGACCTCCATCCTGTGATGG	6169

Db 46331 ACTGGGTCTACTACACTTGGAGCAGGACGTGGAAATGGGACCTCATCTCATG 46390  
 QY 6170 AGAGGGCTGAGCAGCAGCAGATACAGATGTTCCCTGTCTCATGCGAGATTCGGCC 6229  
 Db 46391 AGAGGGCTGAGCAGCAGCAGATACAGATGTTCCCTGTCTCATGCGAGATTCGGCC 46450  
 QY 6230 AGTTTCAAGAGCAGTACAGATCTGTGTGGAAACAAAGTATCAAGCCCTAAGCC 6289  
 Db 46451 AGTTTCAAGAGCAGTACAGATCTGTGTGGAAACAAAGTATCAAGCCCTAAGCC 46510  
 QY 6290 CATTTGGTCTAATTAATCAGAACCCGGGGATGACAGCTGTGAGCAGAGACTTT 6349  
 Db 46511 CATTTGGTCTAATTAATCAGAACCCGGGGATGACAGCTGTGAGCAGAGACTTT 46570  
 QY 6350 TTAATAAGCTCCAGGTGATTTGATCAGCAGCTGGAACAAACAGCTACAGTTCA 6409  
 Db 46571 TTAATAAGCTCCAGGTGATTTGATCAGCAGCTGGAACAAACAGCTACAGTTCA 46630  
 QY 6410 CAGAAAGAGCAAGCTAGGAAAGCTTGGGATGGGAGCTTCTCCAGGCCAGTAGAT 6469  
 Db 46631 CAGAAAGAGCAAGCTAGGAAAGCTTGGGATGGGAGCTTCTCCAGGCCAGTAGAT 46690  
 QY 6470 GGAGGCTGGTATGACAGTGGGAGCTTCTCTGCTGTCATATAGCTATCCATGCACT 6529  
 Db 46691 GGAGGCTGGTATGACAGTGGGAGCTTCTCTGCTGTCATATAGCTATCCATGCACT 46750  
 QY 6530 CATCCATCCATACACCCACCCATCATTTATGACCCATCTCTTCATCCATCATATAC 6589  
 Db 46751 CATCCATCCATACACCCACCCATCATTTATGACCCATCTCTTCATCCATCATATAC 46810  
 QY 6590 CAGCTACCCACCCAGCATCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 6649  
 Db 46811 CAGCTACCCACCCAGCATCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 46870  
 QY 6650 CTTCACCTCATCTTATTCACACAGAGAACTGGTATGTACTAAATGTGGAGATTAAT 6709  
 Db 46871 CTTCACCTCATCTTATTCACACAGAGAACTGGTATGTACTAAATGTGGAGATTAAT 46930  
 QY 6710 TAAATTTTAGAACTCTGTGATTTGATGATGATGATGATGATGATGATGATGATGAT 6769  
 Db 46931 TAAATTTTAGAACTCTGTGATTTGATGATGATGATGATGATGATGATGATGATGAT 46990  
 QY 6770 CAGCAGACGTGTGGCAATGGGAAAGGTTTGGGTTGTTTCTCTCCACCGCTGTG 6829  
 Db 46991 CAGCAGACGTGTGGCAATGGGAAAGGTTTGGGTTGTTTCTCTCCACCGCTGTG 47050  
 QY 6830 GGTTCGGGATTTGAATCAATTTATCGGGCTGGTGGCAAGTCTTTCCACCGAGCA 6889  
 Db 47051 GGTTCGGGATTTGAATCAATTTATCGGGCTGGTGGCAAGTCTTTCCACCGAGCA 47110  
 QY 6890 TTTTGTGACATCATCTTATTTATAGAAAGCATCTTATGTAGTCCAGGCTGGCTCAAGC 6949  
 Db 47111 TTTTGTGACATCATCTTATTTATAGAAAGCATCTTATGTAGTCCAGGCTGGCTCAAGC 47170  
 QY 6950 TTGCTATGTGCGCAGGATGACCTTTAACTCTGCTCTTCCAGCCTCCACCGAGTCTA 7009  
 Db 47171 TTGCTATGTGCGCAGGATGACCTTTAACTCTGCTCTTCCAGCCTCCACCGAGTCTA 47230  
 QY 7010 GGTTCAGGTGTTCAAC----- 7027  
 Db 47231 GGTTCAGGTGTTCAAC----- 47290  
 QY 7028 -----TGTGTAATGCTTTTAAATCCACAGCTGTGGG 7060  
 Db 47291 AGAAATCTGGGCAACAGCTGGGCAATGGTGGTGAATGCCCTTTAATCCACAGCTGTGGG 47350  
 QY 7061 GGGGGGGGGAGAGCGGATCCCTGAGTTGAGGCCAGTTTGTCTACAGATTTTCAGAT 7120  
 Db 47351 GGGGGGGGGAGAGCGGATCCCTGAGTTGAGGCCAGTTTGTCTACAGATTTTCAGAT 47410  
 QY 7121 ACCTGGGCTATACAGGAAACCTTATCCCAACAAACAAACAAACAAACAAATATTT 7180

Db 47411 ACCTGGGCTATACAGGAAACCTTATCCCAACAAACAAACAAACAAACAAATATTT 47470  
 QY 7181 CTGTGCAATATACAGAGATTAGAGATATTAGTAGGGCTGTGAGGAGA 7240  
 Db 47471 CTGTGCAATATACAGAGATTAGAGATATTAGTAGGGCTGTGAGGAGA 47530  
 QY 7241 GTGATGCTTCTTTTGTATTAATTAATTAAGTACTACAGATGCA-----TATC 7292  
 Db 47531 GTGATGCTTCTTTTGTATTAATTAATTAAGTACTACAGATGCA-----TATC 47590  
 QY 7293 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7352  
 Db 47591 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 47650  
 QY 7353 CATCCATCCATATGATGATAGGCCAGGCTGCTTTGACTGTAAATGCTCTCTATTTCTGGG 7412  
 Db 47651 CATCCATCCATATGATGATAGGCCAGGCTGCTTTGACTGTAAATGCTCTCTATTTCTGGG 47710  
 QY 7413 TCACCTTTCACCCCTAGTGTGGTTTACCAACCCAGACATTTATTTATTTTGT 7472  
 Db 47711 TCACCTTTCACCCCTAGTGTGGTTTACCAACCCAGACATTTATTTATTTTGT 47770  
 QY 7473 TATTTATTAATCTAGAGCTCAGGGTGGAGCTCAGGGTCTTGTGATGCTTAAGCAAGCT 7532  
 Db 47771 TATTTATTAATCTAGAGCTCAGGGTGGAGCTCAGGGTCTTGTGATGCTTAAGCAAGCT 47830  
 QY 7533 CTCTGCCACAGAGCTGACCTCAGTCCCATTTTGTTCAGGTGACTGTGACAGTTGT 7592  
 Db 47831 CTCTGCCACAGAGCTGACCTCAGTCCCATTTTGTTCAGGTGACTGTGACAGTTGT 47890  
 QY 7593 CATATTGCGAGGCTATGATGCTCTCTCCACCTCCAGTTCAGACCTTTCTGGTATCC 7652  
 Db 47891 CATATTGCGAGGCTATGATGCTCTCTCCACCTCCAGTTCAGACCTTTCTGGTATCC 47950  
 QY 7653 CAGTGGGGGGCAACTCTGCTCAGCAGAGGCTTCTCCCTGCTTCAGACCTACATAT 7712  
 Db 47951 CAGTGGGGGGCAACTCTGCTCAGCAGAGGCTTCTCCCTGCTTCAGACCTACATAT 48010  
 QY 7713 TTGCTGTCTGAACAGTTGATGTAATGGAGTCCGTTCTGTGATTTCTTTATGGCTGG 7772  
 Db 48011 TTGCTGTCTGAACAGTTGATGTAATGGAGTCCGTTCTGTGATTTCTTTATGGCTGG 48070  
 QY 7773 CCCCTTATCTTAGCAGATTTGTGTGGGCAATGTGTACATCTATCTATCTATC 7832  
 Db 48071 CCCCTTATCTTAGCAGATTTGTGTGGGCAATGTGTACATCTATCTATCTATC 48130  
 QY 7833 ATCATCTATAGGCTTATAGTGTCTTGTGTGGATTAACACATTTCTGTTTCAATTTAC 7892  
 Db 48131 ATCATCTATAGGCTTATAGTGTCTTGTGTGGATTAACACATTTCTGTTTCAATTTAC 48190  
 QY 7893 TGAATGAAATTTGTGGCCACCACCACCTTTTATTTTATTTGAGCAAGGCTTTTC 7952  
 Db 48191 TGAATGAAATTTGTGGCCACCACCACCTTTTATTTTATTTGAGCAAGGCTTTTC 48250  
 QY 7953 TGTGTAATCTTGAATCTTGGCTGTCTGAGCTCATCTGTAGACAGGCTGTGAGGCTG 8012  
 Db 48251 TGTGTAATCTTGAATCTTGGCTGTCTGAGCTCATCTGTAGACAGGCTGTGAGGCTG 48310  
 QY 8013 TCCCTCAGCTTTTGAACCTCTGTGAACAGAGTAGCATGAACTTCAAGACAAATTTTCT 8072  
 Db 48311 TCCCTCAGCTTTTGAACCTCTGTGAACAGAGTAGCATGAACTTCAAGACAAATTTTCT 48370  
 QY 8073 GTTTTGGTTTGTATTTTACATTTGTTGTGTATGCTGTATATGTGCAATGTTGTGCTTT 8132  
 Db 48371 GTTTTGGTTTGTATTTTACATTTGTTGTGTATGCTGTATATGTGCAATGTTGTGCTTT 48430  
 QY 8133 CAGGTGCTACATGTGTGACTGTGTGTGGCAGAGAAACAAACCAAGTGGCATTTCT 8192  
 Db 48431 CAGGTGCTACATGTGTGACTGTGTGTGGCAGAGAAACCAAGTGGCATTTCT 48490  
 QY 8193 CAGATCTACAGATCTGTTAATATGATATGATATGATATGATATGATATGATATGAT 8252  
 Db 48491 CAGATCTACAGATCTGTTAATATGATATGATATGATATGATATGATATGATATGAT 48550

```

OY 8253 TATGAGGTATTTTGGAGTTTTCACCTTCCCTTGTGGGCTCCGCATTAACTCAG 8312
    |||||||
DB 48551 TATGAGGTATTTTGGAGTTTTCACCTTCCCTTGTGGGCTCCGCATTAACTCAG 48610
OY 8313 CTCTCGGGGCTAGTGAAGCAATGCTTCACTCGATGAGCCATCTGCTCCCTGCTGCCA 8372
    |||||||
DB 48611 CTCTCGGGGCTAGTGAAGCAATGCTTCACTCGATGAGCCATCTGCTCCCTGCTGCCA 48670
OY 8373 CCTCTCTTATTTTCCCGATGAGGCTAGCGACTGCTGCTTAAGCTCACCAGTCA 8432
    |||||||
DB 48671 CTCTCTCTTATTTTCCCGATGAGGCTAGCGACTGCTGCTTAAGCTCACCAGTCA 48730
OY 8433 TCCAGAGTGGCTAGCGAGGAGACTCAGGAGATATGCTGCTCCAGAGTCTAG 8492
    |||||||
DB 48731 TCCAGAGTGGCTAGCGAGGAGACTCAGGAGATATGCTGCTCCAGAGTCTAG 48790
OY 8493 AATTACAGGCATACATCAGTCTGTGAAGATTTTAACTGAATCCTGAGATAGAGCAG 8552
    |||||||
DB 48791 AATTACAGGCATACATCAGTCTGTGAAGATTTTAACTGAATCCTGAGATAGAGCAG 48850
OY 8553 CACCTACCATGAGGCTTCTTTTGTGTTGCTTGTGCTTCCCTGCTAATAGATCAG 8612
    |||||||
DB 48851 CACCTACCATGAGGCTTCTTTTGTGTTGCTTGTGCTTCCCTGCTAATAGATCAG 48910
OY 8613 CAGTCTGAATAGTGTAGCTGCGCTACATTAACATCTGTCTCAAAAAGCCTATAGAGT 8672
    |||||||
DB 48911 CAGTCTGAATAGTGTAGCTGCGCTACATTAACATCTGTCTCAAAAAGCCTATAGAGT 48970
OY 8673 AGGAGGTGAGGCTAAAGAAAGAGCCTTAAGCCGGCTGTGATAGCACAGATAGCCTG 8732
    |||||||
DB 48971 AGGAGGTGAGGCTAAAGAAAGAGCCTTAAGCCGGCTGTGATAGCACAGATAGCCTG 49030
OY 8733 CACATATATGAAGACCTTTTCAAAAACATGAGGAGGCTATGTTTAAAGTCTCG 8792
    |||||||
DB 49031 CACATATATGAAGACCTTTTCAAAAACATGAGGAGGCTATGTTTAAAGTCTCG 49090
OY 8793 GCTGTGTAAAGGCACTTAAAGGAGGCTATGATTAATGCTAAAGAAAGTCAATC 8852
    |||||||
DB 49091 GCTGTGTAAAGGCACTTAAAGGAGGCTATGATTAATGCTAAAGAAAGTCAATC 49150
OY 8853 AAAGCCGGGTGGGCTAGAGTGTGACTACAGTGTCAAGACCCCATAGAGGCCA 8912
    |||||||
DB 49151 AAAGCCGGGTGGGCTAGAGTGTGACTACAGTGTCAAGACCCCATAGAGGCCA 49210
OY 8913 GTTTCCTCTTCTCCGCTGAGGCTCAAGCTGCTGAGGCTGCTCAACATGCTT 8972
    |||||||
DB 49211 GTTTCCTCTTCTCCGCTGAGGCTCAAGCTGCTGAGGCTGCTCAACATGCTT 49270
OY 8973 CTCTCTAGGCTGTCCACATG 8995
    |||||||
DB 49271 CTCTCTAGGCTGTCCACATG 49293

```

```

RESULT 3
AC132885
LOCUS AC132885 170074 bp DNA linear HTG 12-MAR-2003
DEFINITION Mus musculus clone RP24-263015, WORKING DRAFT SEQUENCE, 9 unordered
          pieces
ACCESSION AC132885
VERSION AC132885.3 GI:28927760
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 170074)
AUTHORS Birren,B., Nusbaum,C., and Lander,E.
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170074)
AUTHORS Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
          Batra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,  
Karatas,A., Kells,C., Lander,S., Lander,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlengv,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Plerie,N., Raymond,C., Retla,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 170074)

Birren,B., Nusbaum,C., Lander,E., Abouelkell,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Batra,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Boukhgalter,B., Camata,J., Chang,J., Choepel,Y.,  
Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hager,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lander,S., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,D., Matthews,C., McCarthy,M.,  
Meldrum,J., Menus,L., Mihova,T., Mlengv,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Plerie,N.,  
Rachupka,A., Ramasamy,V., Raymond,C., Retla,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 12, 2003 this sequence version RepeatMasker:  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L27142  
Center clone name: 263\_O\_15

----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 168429 bases at least Q40  
Consensus quality: 168822 bases at least Q30  
Consensus quality: 169112 bases at least Q20  
Insert size: 164000; agarose-fp  
Insert size: 169274; sum-of-ctnfigs  
Quality coverage: 8.7 in Q20 bases; agarose-fp  
Quality coverage: 8.4 in Q20 bases; sum-of-ctnfigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence



QY	4093	AGTTTGGACCTACTGACACCAAGTATATCTGCTCAGAAATCCCTGTCACTTGAGGTT	4152
Db	1379	AGTTTGGGACTACTACACAGCCAAAGGTATCTGCTCAGAAATCCCTGTCACTTGAGGTT	1438
QY	4153	GGGAGAAATCTGCTCTGGGGGCTTCAGGTCCTTGGTTGGACGAGGATACCTTGTATTA	4212
Db	1439	GGGAGAAATCTGCTCTGGGGGCTTCAGGTCCTTGGTTGGACGAGGATACCTTGTATTA	1498
QY	4213	GGGAGTAAGTATGTTATGTTTACTATCTCTGTCAGTAAACCTGGAACTGATAAA	4272
Db	1499	GGGAGTAAGTATGTTATGTTTACTATCTCTGTCAGTAAACCTGGAACTGATAAA	1558
QY	4273	CCCAGGCAAGCCGCCAGAGATTTCTTACAGTTGTATACCCAGACAACAAGACATGATTA	4332
Db	1559	CCCAGGCAAGCCGCCAGAGATTTCTTACAGTTGTATACCCAGACAACAAGACATGATTA	1618
QY	4333	TGCAAGATATAGTACTGCTGGGAGAGAAGAACTTAAACCCCCCAAGAGCCACAGGATC	4392
Db	1619	TGCAAGATATAGTACTGCTGGGAGAGAAGAACTTAAACCCCCCAAGAGCCACAGGATC	1678
QY	4393	CGTTCCCTAGTTCAACATGCCAGATATGATGCTACTATAGGCTGTGATTTGGTAGC	4452
Db	1679	CGTTCCCTAGTTCAACATGCCAGATATGATGCTACTATAGGCTGTGATTTGGTAGC	1738
QY	4453	TACAAAGCATAGTATGTTTCAATGCTGTAGTGTATATATCTGAGCACTTGGAGGCTGA	4512
Db	1739	TACAAAGCATAGTATGTTTCAATGCTGTAGTGTATATCTGAGCACTTGGAGGCTGA	1798
QY	4513	AGCAGAGAGATTTGCTATATGTTTGAAGCCAGCCTGAGTATAGACGAGCACTTGTCTTT	4572
Db	1799	AGCAGAGAGATTTGCTATATGTTTGAAGCCAGCCTGAGTATATAGACGAGCACTTGTCTTT	1858
QY	4573	AAGAAAAATATGAAAGCCCGACAGTGTGTGGACACGCCTTTATCTCCAGCACTTGGAGG	4632
Db	1859	AAGAAAAATATGAAAGCCCGACAGTGTGTGGACACGCCTTTATCTCCAGCACTTGGAGG	1918
QY	4633	CAGAAGCAGGAGATTTCTGAGATTCAGAGCCAGCCTGCTATATAGAGAGTCCAGAGAC	4692
Db	1919	CAGAAGCAGGAGATTTCTGAGATTCAGAGCCAGCCTGCTATATAGAGAGTCCAGAGAC	1978
QY	4693	AGCCAGGCTTACACAGAGAAACCTGTTTGAAGAAAAACGAGAAAAACAAACAAACAAA	4752
Db	1979	AGCCAGGCTTACACAGAGAAACCTGTTTGAAGAAAAACGAGAAAAACAAACAAACAAA	2038
QY	4753	CAAAACAAACCCAAACCCAAACCTCTCATCTCTCATCTCTCTAGGCTGTGCT	4812
Db	2039	CAAAACAAACCCAAACCCAAACCTCTCATCTCTCATCTCTCTAGGCTGTGCT	2098
QY	4813	GTCATGCTGATAGACTTTGGGAGCTTACAGATTTATATATAAATAGGCTTTTATCACTG	4872
Db	2099	GTCATGCTGATAGACTTTGGGAGCTTACAGATTTATATATAAATAGGCTTTTATCACTG	2158
QY	4873	GTCAGAGACGAGAAAGGTTTCACTGCTGGAGACACACTGGGACCTTGAGAAAGTACTCTTG	4932
Db	2159	GTCAGAGACGAGAAAGGTTTCACTGCTGGAGACACACTGGGACCTTGAGAAAGTACTCTTG	2218
QY	4933	CCAGCCCAAAATTTGTGGGAAGGCTCTTGAGAGGAAGTGTCCCGA	4979
Db	2219	CCAGCCCAAAATTTGTGGGAAGGCTCTTGAGAGGAAGTGTCCCGA	2278
QY	4980	-----TCAG	4983
Db	2279	AATATGAAGAAGTGGGAAAGCAAGGCAAGAGCAGCTGGGTGTAAAGGACGACAGTACG	2338
QY	4984	ACTACTGTTCTTGAAGGCAAGAGAGAGGCTTGGAAAGATGTTGGTGACAGACAGTTGGA	5043
Db	2339	ACTACTGTTCTTGAAGGCAAGAGAGAGGCTTGGAAAGATGTTGGTGACAGACAGTTGGA	2398
QY	5044	ACAGAAAGACAGGAGGGGAGGAGCATCCAAAGATTCATGAAATGATAGTACTTGGTGTCT	5103
Db	2399	ACAGAAAGACAGGAGGGGAGGAGCATCCAAAGATTCATGAAATGATAGTACTTGGTGTCT	2458



|||||  
Db 3531 CACCAGGTACAAAGTTCCTTGTCATGACGAGGATTCCTGGCCAGTTTCAAGGAC 3590  
QY 6244 TAAGAGCTCATCTCTGGTGGAAACAAAGTATCCAAAGCCCTTAAGCCCATTTGGTCTAAT 6303  
Db 3591 TAAGAGCTCATCTCTGGTGGAAACAAAGTATCCAAAGCCCTTAAGCCCATTTGGTCTAAT 3650  
QY 6304 TAAATCAGAAACCCCTGGGGATGCAAGGCTCTGAGCAGCAGAGAGCTTTTAAAAAGCTCCA 6363  
Db 3651 TAAATCAGAAACCCCTGGGGATGCAAGGCTCTGAGCAGCAGAGAGCTTTTAAAAAGCTCCA 3710  
QY 6364 GGTGATTCGTATCAGCAGCTGGAACAAACAGCTACAGTTCAACAGAAAGAGGAAA 6423  
Db 3711 GGTGATTCGTATCAGCAGCTGGAACAAACAGCTACAGTTCAACAGAAAGAGGAAA 3770  
QY 6424 GCTAGGAAAGCTTGGGATGGGAGGCTTCTCCAGGCGAGTATGAGAGGCTGTATAGC 6483  
Db 3771 GCTAGGAAAGCTTGGGATGGGAGGCTTCTCCAGGCGAGTATGAGAGGCTGTATAGC 3830  
QY 6484 AGTGTGGCAGGCTTCTCTGCTGTCAATATAGCTATCCATCCATCATCCATACATA 6543  
Db 3831 AGTGTGGCAGGCTTCTCTGCTGTCAATATAGCTATCCATCCATCATCCATACATA 3890  
QY 6544 CCCACCATCCATTTATGACCCATCCTTCATCCATCCATCCATCCATCCATCCATCCATCCAT 6603  
Db 3891 CCCACCATCCATTTATGACCCATCCTTCATCCATCCATCCATCCATCCATCCATCCATCCAT 3950  
QY 6604 CGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6663  
Db 3951 CGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4010  
QY 6664 TTATCCAGAGAACTGGTATGTTACTAAATGTGGAGATTTAATTTTATAGAGC 6723  
Db 4011 TTATCCAGAGAACTGGTATGTTACTAAATGTGGAGATTTAATTTTATAGAGC 4070  
QY 6724 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6783  
Db 4071 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4130  
QY 6784 CAATCGAGAAAGTTTGGGTGTGTTTCTCTCCACCGTGGGTTCTGGGATG 6843  
Db 4131 CAATCGAGAAAGTTTGGGTGTGTTTCTCTCCACCGTGGGTTCTGGGATG 4190  
QY 6844 AACTCAATTTATGGGCTGTGGCAAGTCTTTACACGAGCCATTTTGTGACACAT 6903  
Db 4191 AACTCAATTTATGGGCTGTGGCAAGTCTTTACACGAGCCATTTTGTGACACAT 4250  
QY 6904 CATTATTTATAGAAAGCATTTATGATGTCAGGCTGGCTCAAGCTTGTATGTGCCA 6963  
Db 4251 CATTATTTATAGAAAGCATTTATGATGTCAGGCTGGCTCAAGCTTGTATGTGCCA 4310  
QY 6964 CGGATGACCTTTAACTCTGCTCTCTCCAGCCTCCACCGAGTCTAGGTTTACAGGTGT 7023  
Db 4311 CGGATGACCTTTAACTCTGCTCTCTCCAGCCTCCACCGAGTCTAGGTTTACAGGTGT 4370  
QY 7024 CAAC----- 7027  
Db 4371 CAACACCACACCAATTTAATGCTCAGAGATTTTGTACTGAGCAGGCAAAATCTGGGCAA 4430  
QY 7028 -----TGGTAATGCTTTAATCCAGACCTGTGGGGGGGGGGAGG 7074  
Db 4431 CAGCTGGGCAATGGTGGATGCTTTAATCCAGACCTGTGGGGGGGGGGAGG 4490  
QY 7075 CGGATCCCTGATTTGGAGGCAAGTGGTCTACAGAGTTTCAAGATTAACCTGGGGCTATAC 7134  
Db 4491 CGGATCCCTGATTTGGAGGCAAGTGGTCTACAGAGTTTCAAGATTAACCTGGGGCTATAC 4550  
QY 7135 AGGGAACCTATCCCAAAACAAACAAACAAACAAACAAATTTCTGTGCAATATCA 7194  
Db 4551 AGGGAACCTATCCCAAAACAAACAAACAAACAAACAAATTTCTGTGCAATATCA 4610  
QY 7195 CAGAGATTAGAGATTTAGTAGGATGATGGGCTGGTGGAGAGATCATGCTTCTTT 7254  
|||||

Db 4611 CAGAGATTAGAGATTTAGTAGGATGATGGGCTGGTGGAGAGATCATGCTTCTTT 4670  
QY 7255 TGTATTTAATATAGTAAGTACTCACAAGATGAT-----TATCATATCATATCA 7306  
Db 4671 TGTATTTAATATAGTAAGTACTCACAAGATGATTTATCATATCATATCATATCA 4730  
QY 7307 TCTATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATCA 7366  
Db 4731 TCTATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATCA 4790  
QY 7367 TCGATATGAGCCAGGCTGCTTGAATGAGTCCATTTCTGGGCAACCTTCACCC 7426  
Db 4791 TCGATATGAGCCAGGCTGCTTGAATGAGTCCATTTCTGGGCAACCTTCACCC 4850  
QY 7427 CTAGTGTGGGTTTACCAACACCAGACATTTATTTATTTTATTTATTTATTTATCT 7486  
Db 4851 CTAGTGTGGGTTTACCAACACCAGACATTTATTTATTTTATTTATTTATTTATCT 4910  
QY 7487 AGGAGCTCAGAGGTGGGACTCAGGCTCTTGATGCTAAGCAAGCTCTGCCAGAGC 7546  
Db 4911 AGGAGCTCAGAGGTGGGACTCAGGCTCTTGATGCTAAGCAAGCTCTGCCAGAGC 4970  
QY 7547 TGCAGCTCCAGTCCCATTTTGTTCAGGTGACTCTGTGACAGTTGCAATTTGCCAGCG 7606  
Db 4971 TGCAGCTCCAGTCCCATTTTGTTCAGGTGACTCTGTGACAGTTGCAATTTGCCAGCG 5030  
QY 7607 TATGATGCTCTCCACCTCCAGTTCACAGACTTCTGTGCTATCCAGTGGGCGCAA 7666  
Db 5031 TATGATGCTCTCCACCTCCAGTTCACAGACTTCTGTGCTATCCAGTGGGCGCAA 5090  
QY 7667 CTCTGTGCTCACCAGTCCCTGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7726  
Db 5091 CTCTGTGCTCACCAGTCCCTGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5150  
QY 7727 AGTTCAATGTAATGGGATGGGCTTCTGTGATTTCTTTATGGCTGGCCCTTATCTTAG 7786  
Db 5151 AGTTCAATGTAATGGGATGGGCTTCTGTGATTTCTTTATGGCTGGCCCTTATCTTAG 5210  
QY 7787 CACAGTTTGTGTTGGGCATGTGTCACTGTATCTATCTATCTATCTATCTATCTATCTAT 7846  
Db 5211 CACAGTTTGTGTTGGGCATGTGTCACTGTATCTATCTATCTATCTATCTATCTATCTAT 5270  
QY 7847 TAATAGTTCCTTTGTGTGATTAACCACTTCTGTTCATTTACTGATGGAATTTGT 7906  
Db 5271 TAATAGTTCCTTTGTGTGATTAACCACTTCTGTTCATTTACTGATGGAATTTGT 5330  
QY 7907 GGCCCAACCCCAACCTTTTCTTTTATTTAGAGCAAGGCTCTTCTGTGTAATCTTGA 7966  
Db 5331 GGCCCAACCCCAACCTTTTCTTTTATTTAGAGCAAGGCTCTTCTGTGTAATCTTGA 5390  
QY 7967 ATCTTGGCTGTCTGAGCTCAGTCTGTAGCAAGGCTGTGAGGCTGCTCCACATTTTG 8026  
Db 5391 ATCTTGGCTGTCTGAGCTCAGTCTGTAGCAAGGCTGTGAGGCTGCTCCACATTTTG 5450  
QY 8027 ACACCTCTGTGAACAGATGACATGAACCTTCAACAACTTTCTGTGTTGGTGTGTT 8086  
Db 5451 ACACCTCTGTGAACAGATGACATGAACCTTCAACAACTTTCTGTGTTGGTGTGTT 5510  
QY 8087 TTTACATTTTGTGTGTATGCGTATATGATGATGATTTGTGCTTACAGTGTCCACATG 8146  
Db 5511 TTTACATTTTGTGTGTATGCGTATATGATGATGATTTGTGCTTACAGTGTCCACATG 5570  
QY 8147 TGTGTACCTGTGTGGGACAGAAACAAACGATGTCATTTCTCAGATTAACGAT 8206  
Db 5571 TGTGTACCTGTGTGGGACAGAAACAAACGATGTCATTTCTCAGATTAACGAT 5630  
QY 8207 CTTGTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8266  
Db 5631 CTTGTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5690  
QY 8267 GTTGGAGTTTTCACCTTCTCCCTTGTGGGCTCTCCGATTTAAACCTCAGCTCCCTGGGCTAGT 8326  
Db 5691 GTTGGAGTTTTCACCTTCTCCCTTGTGGGCTCTCCGATTTAAACCTCAGCTCCCTGGGCTAGT 5750  
|||||



OY		8327	GAGCAATGCCCTTCACTCGATGATGAGGCATTCGTGCTCCCTGGTGACACTCCTCCTTAATT	8386
Db		5751	GAGCAATGCCCTTCACTCGATGATGAGGCATTCGTGCTCCCTGGTGACACTCCTCCTTAATT	5810
OY		8387	CCCAGATGGGACTACGCACTGCTACCTGGGCTTAAAGCTCACAAAGTCATCCAGTAGGCGTAG	8446
Db		5811	CCCAGATGGGACTACGCACTGCTACCTGGGCTTAAAGCTCACAAAGTCATCCAGTAGGCGTAG	5870
OY		8447	CCAGGGAGACTCAGGGATATGCTGGCCTCTGCTCCACAGTGTGAATTTACAGGCATAC	8506
Db		5871	CCAGGGAGACTCAGGGATATGCTGGCCTCTGCTCCACAGTGTGAATTTACAGGCATAC	5930
OY		8507	ATCATGCTGGAAAGTTTTTTAAACMGAAATCCTGGAGGATAGGCGAGGCACTCTACCATAAG	8566
Db		5931	ATCATGCTGGAAAGTTTTTTAAACMGAAATCCTGGAGGATAGGCGAGGCACTCTACCATAAG	5990
OY		8567	AGGGTTCCTTTTGTGTGGTTGGTTGGTTCCCTCTCATTAAGATTCAGGCAAGTCTGAAATAGT	8626
Db		5991	AGGGTTCCTTTTGTGTGGTTGGTTGGTTCCCTCTCATTAAGATTCAGGCAAGTCTGAAATAGT	6050
OY		8627	GTACCCCTGGGCTAACAATAACATCTTGTCTCAA AAAAGCCTATATAGAGGTAGGAGGTGAGGC	8686
Db		6051	GTACCCCTGGGCTAACAATAACATCTTGTCTCAA AAAAGCCTATATAGAGGTAGGAGGTGAGGC	6110
OY		8687	TAAAGAAGAGCCTTAAAGCCGGCTGTGATAGACACAGAGATATAGCCTGTATATAGAACAG	8746
Db		6111	TAAAGAAGAGCCTTAAAGCCGGCTGTGATAGACACAGAGATATAGCCTGTATATAGAACAG	6170
OY		8747	ACCTTGTTTCAAAAACATGAGGAGGAGGGGTATGTTTTAAAGTCTGGGGCTGTGTAAACAGGC	8806
Db		6171	ACCTTGTTTCAAAAACATGAGGAGGAGGGGTATGTTTTAAAGTCTGGGGCTGTGTAAACAGGC	6230
OY		8807	ACTAAGGAGGCCATGTATAGACATTTTGACTAAGAAAGATTCATTCATCAAAGCCGGGTGGGC	8866
Db		6231	ACTAAGGAGGCCATGTATAGACATTTTGACTAAGAAAGATTCATTCATCAAAGCCGGGTGGGC	6290
OY		8867	AGGGTAGAGGTGGACTACAGTGGTCAAGACCCCATAGAGAAAGCCAGTTTCCCTTCTTCCG	8926
Db		6291	AGGGTAGAGGTGGACTACAGTGGTCAAGACCCCATAGAGAAAGCCAGTTTCCCTTCTTCCG	6350
OY		8927	TCTGGGCCCTCAAGCCTGGCTCGACGGCCACTGCTCTCAACATGCTTCTCTTAGGCTCG	8986
Db		6351	TCTGGGCCCTCAAGCCTGGCTCGACGGCCACTGCTCTCAACATGCTTCTCTTAGGCTCG	6410
OY		8987	TCCACCATG 8995	
Db		6411	TCCACCATG 6419	
RESULT 4 LOCUS AC115173				
DEFINITION	Rattus norvegicus clone CH230-286017, WORKING DRAFT SEQUENCE, 3			
ACCESSION	AC115173			
VERSION	AC115173.4 GI:25006753			
KEYWORDS	HTGS; HTGS; PHASE1; HTGS; DRAFT; HTGS_FULLUTOP.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
REFERENCE	1 (bases 1 to 193472)			
AUTHORS	Muzny,D.,Marie,M.,Metzker,M.,Lee,A.,Abramzon,S.,Adams,C.,Alder,J.,Allen,C.,Allen,H.,Alsbrooks,S.,Amis,A.,Angiano,D.,Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,Baldwin,D.,Bandaranaike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,Bliswalo,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,Bryant,N.,Buhey,C.,Burch,P.,Burrell,K.,Calderon,E.,Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,			

TITLE	JOURNAL
REFERENCE	AUTHORS
TITLE	JOURNAL
Del Valle, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederlich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Gazda, M., Gebregorgis, E., Geer, K., Gill, R., Gillingham, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, V., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hunne, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Izu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Meshawari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., McWhiney, S., McLeod, M.P., McNelli, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Molja, E., Montemayor, J., Moore, S., Morgan, A., Morris, K., Morris, S., Munadasa, M., Murphy, M., Naif, L., Nankevits, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nockelmech, O., Okwunodu, G., Olarpungazon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polndexter, A., Popovic, D., Prlums, E., Pu, L., L., Piazzi, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shwartsbeyn, A., Sisoon, I., Sitter, C.D., Smajz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabot, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Velas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weidhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.	Unpublished Direct Submission 2 (bases 1 to 193472)
Morley, R.C.	Direct Submission Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 193472)
Rat Genome Sequencing Consortium.	Direct Submission Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Nov 15, 2002 this sequence version replaced g1:23678661. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/atlas/">http://www.hgsc.bcm.tmc.edu/projects/atlas/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information

```
Center project name: GOND
Center clone name: CH230-286017
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 185211 bases at least Q40
Consensus quality: 186794 bases at least Q30
Consensus quality: 188105 bases at least Q20
Estimated insert size: 184863; sum-of-contrigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bjsc.bcm.tmc.edu/docs/genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 190920: contig of 190920 bp in length
* 190921 191020: gap of unknown length
* 191021 192270: contig of 1250 bp in length
* 192271 192370: gap of unknown length
* 192371 193472: contig of 1102 bp in length.
FEATURES
Source
1. 193472
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-286017"
1. 1118
/feature="wgs_end_extension"
clone_end:T7
/feature="wgs_end_extension"
clone_end:T7
/feature="clone_boundary"
site:
end_sequence:RXAHM93TJC"
4668..5345
/feature="clone_boundary"
clone_end:T7
site:
end_sequence:RXAHM93TJB"
185119..186013
/feature="clone_boundary"
clone_end:Sp6
site:
end_sequence:RXAHM93TJB"
187253..188705
/feature="wgs_end_extension"
clone_end:T7"
189469..190920
/feature="wgs_end_extension"
clone_end:T7"
BASE COUNT 51460 a 44772 c 44401 g 48651 t 4188 others
ORIGIN
Query Match 34.7%; Score 3116.8; DB 2; Length 193472;
Best Local Similarity 67.7%; Pred. No. 0;
Matches 6694; Conservative 0; Mismatches 1525; Indels 1662; Gaps 102;
QY 93 GGTTTAGCCAGCTTACGCTACATGAACCTTTGTT-----TGTGTTGTTGTTGTTGTT 143
DB 119505 GCTTTGGCCAGCTTGCGCTACATGAACCTTTGTTGTTGTTGTTGTTGTTGTTGTTG 119564
QY 144 TTTAAAGCATTAATAATATCCATGAAGAGGTGGCAG-TGCTGGCAGACACCTTTAA 202
DB 119565 TTTTAAGCATTAATAATATCCATGAAGAGGTGGCTGGTGTCTGACACACCTTTAA 119624
QY 203 TTCACATATTGAGAGCAGAGCAGAGCATCTGTGAGTTGCAAGTCAGAGCTTACT 262
DB 119625 TTCACAGATTTCAGAGAGGTGAAGACAGCAGAGCTCTGTGAGTTGCAAGCAGCCTGTCT 119684
```

```
QY 263 GCAGAGCTAGTCCAGAGATGGCAGAGGCTACACAGAGAAACCTTGTCATTAACCAAA 322
DB 119685 ACAGAGTAGTTCAGAGACAGCAGAGGTTACACAGAGAAACCTGTCTAGAGAAACCAA 119744
QY 323 GTGTGCTGTACTACTA----- 340
DB 119745 GTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 119804
QY 341 -----ATGCCATGAGAGAAATTTGAGTCCATTCAGATGG-ACCATC 381
DB 119805 GCAGCAGCAGCAGCATAGAGAAATTTGGGGTCTATTGGGTCATTCATGATGAAACATC 119864
QY 382 CTATAGATGATTCCTTGACCCAGGTAAAGTAACTATGATGGGAAAGGGATGGACTG 441
DB 119865 CTATAGAGATTCCTTGACCCAGGTAAAGTAACTATGATGGGAAAGGGATGGACTG 119924
QY 442 TCCATATTAAAGAGCGAGGCGATGGCTATTCATTCATTCATTCATTCATTCATTCATTC 501
DB 119925 CCTAGATTAAAGAGCGAGGCGATGGCTATTCATTCATTCATTCATTCATTCATTCATTC 119984
QY 502 TGATTAAGGCCCAAGAGAGTGAAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 561
DB 119985 TGATTAAGTCCAAAGTGAAGTGAATATCATGTGAAGTGGAGTGGAGTGGAGTGG 120042
QY 562 ACATGAGCACTTATAACACTTATAACACTGCGCAGCGCTTCAGGTTGAAGATCACT 621
DB 120043 -----GGTTTATTAATATCTGCGACAGACCCCTCAGATATGAAGATCACT 120086
QY 622 TTCAACCAAGAGAGAGAGTGGCTGCGTCGTCAGGATGGCGAGCTGGCTGACGA 681
DB 120087 TTCAACCAAGAGAGAGAGTGGCTGCGTCGTCAGGATGGCGAGCTGGCTGACGA 120142
QY 682 AGAGTGAATATTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 741
DB 120143 ACAGGCTCGTTAATGAAGTGGCTGCGTCGTCAGGATGGCGAGCTGGCTGACGA 120202
QY 742 CAATATGCTTAACCTCCCTAGTCCAGAGATGGCTGTTACACTGTTCTGCTTCCATCT 801
DB 120203 CAATATGCTTAACCTCCCTAGTCCAGAGATGGCTGTTACACTGTTCTGCTTCCAT 120262
QY 802 GGTTCATTTTGCACAGCAGAGAAATTTAGAAATGTGGATTTATTTGTGGTGGAGC 861
DB 120263 GGTTCATTTTGCACAGCAGAGAAATTTAGAAATGTGGATTTATTTGTGGTGGAG 120322
QY 862 ACCATCAGAGGCTTTTACATTTTACAGCAGCATGTTTACTTAACCTGGCTACTTCTCA 921
DB 120323 AACATCAGAGGCTTTTACATTTTACAGCAGCATGTTTACTTAACCTGGCTACTTCT 120381
QY 922 GGTTCATTTTGCACAGCAGAGAAATTTAGAAATGTGGATTTATTTGTGGTGGAGC 979
DB 120382 CAATTAAGCAGAAATTTTGCATTAATTTGATTTTGTGGCTGAGTGGAGTCAAT 120441
QY 980 ACGTATGATTAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1031
DB 120442 ATGATGATGATTAAGAGTATGATGATGATGATGATGATGATGATGATGATGAT 120501
QY 1032 TCCCGCAGAGTGAAGTCAACGAGCGTTGATGATGATGATGATGATGATGATGATGATG 1091
DB 120502 TCCCGCAGAGTGAAGTCAACGAGCGTTGATGATGATGATGATGATGATGATGATGATG 120561
QY 1092 TGGGTTCTGTGCAAGAGCAGAGTGGCTTTAAACATGGGACACAGCTCTAGGCTTA 1151
DB 120562 TGGGTTCTGTGCAAGAGCAGAGTGGCTTTAAACATGGGACACAGCTCTAGGCTTA 120621
QY 1152 GTATATCTTATGTTTAAATATATATATATATATATATATATATATATATATATATAT 1211
DB 120622 ACATATTTTAAATTTTAAATATATATATATATATATATATATATATATATATATAT 120681
QY 1212 ATCCAGCAGCTTGAGAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1271
DB 120682 AGTGTCTGTGAGAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 120741
```

QY	1272	ACCTTG6CCCTGTTTTTTTTTGTGTTTTTCTTTATATGACACTGGTGTCTTACCTGCGTAT	1331
Db	120742	AAGAAAAAATATATATATATCTCTGTGGACTGTGACATAGCCTAATTTCCAGCAC	120801
QY	1332	GT-----	1333
Db	120802	TTGAGAGGCTGGGGGTGGGATCATAAATTCAGTTGGGGTAAAGATGTGGCCCTGTT	120861
QY	1334	-----CCGTGCAGGGGTGCAGATCCCTTGGAGCGTGAAGTTAAAGACAGTTGTGATCAC	1387
Db	120862	TTCCCCCTTATATGTCACACTGGTGTCTTGCTGTGTGTGTGAGGAAATTCGATCCCTG	120921
QY	1388	GCTGCCGTTACAGATGCTGGAAATTAAGACCCAGGTGTCCCTAGAGAGACGCCAGTGTCTC	1447
Db	120922	GAGCTGCCATATGGGTGCTGGAATTAAGACCCAGACACCCCTGGAAAGAGCCCAATGCTC	120981
QY	1448	TTAACTTTCGAGCCACCCCTCCAACTGCTTTTATAGACTCTTAACCTTTTGTGTAAAG	1507
Db	120982	TTAACTTTCGAGCCACCCCTCCAGCCCACTTAAGAAATTCCTAACCCTTTTGTGTGTGT	121041
QY	1508	TGGGAACCTAGT-----	1544
Db	121042	GTGTGTGTGTGTGTGAATATGGAATTTGAACCTTAGAGGCTTGTACACCTAATGTGTGTG	121101
QY	1545	CTGGCCTGTAGCATCTAGTAGGCCCTGACCCACACGACTAGTGATACAGTTTAAGGGCAA	1604
Db	121102	CTGTCTGAAATATCACTAGAGCCACACCCACGCTGATGATGATATACCTAATGCGAA	121161
QY	1605	ACACTTTAACATGACAAATAGTTGGATAGATGAATTA--TATCTGTAGCTAATTTGGTTA	1662
Db	121162	ACATTTAGCAAAAGACAAAGATTTTGGGTAGATTTGAAGACCCAGTCTGTAGCTAATGGTTA	121221
QY	1663	GGGTACCTTCTGCTGCTTACATATGTGTGTGAGAGATGAAAAATAGACCTTGAGT	1722
Db	121222	GGGTACCTTCTGCTGCTTACCTTACGATGTGTGTGAGAGATGAAAAATAGACCTTGAGT	121275
QY	1723	CTAGTCTTGGAACCCACAGAGGCGAGCGAAGAACCACTCTGTAAGTGTCTCTGAGCT	1782
Db	121276	TGAGTCCCGGACCTGACATGTGTGGGTGGAAGCCACCTCTGAAGTTG--TCTGTGACCT	121334
QY	1783	TCACATACACCTTTCACATATATAGTTACATGATTAATTAATTAATTAATTTCTTTTAAA	1842
Db	121335	CTGCATACAAAC---CGTAAGAGTTATATATGACAATGAATTAATTAATTAATTAATTAAGAA	121390
QY	1843	GGTATATGTTGGGAGGAGAGATGGGTGCTGCCAGAGAGCACTGCTGCTGTGAGAG	1902
Db	121391	GGTATAT--GTTGGGCGAGAGAGATGGGTGCTGCCAGAGAGCACTGCTGCTTCTT---AAA	121446
QY	1903	GACCTAGATTCACTTCCAGGACTCATATGTGGGCTCACAGGCACTCTGTAAATCCAGTTC	1962
Db	121447	GCCCTAGATTCACTTCCAGGACTCATATGTGGGCTCACAGGCACTCTGTAAATCCAGTTC	121505
QY	1963	CAGAGGGTTCACACCCCTCTTGGGCTTCACAGGCACTCATATAGTATGACACAGAT	2022
Db	121506	CAGAGGGTTCACACCCCTCTTGGGCTTCGCCAGGCACTCATATAGTATGACACAGAT	121565
QY	2023	ACATGCAAGCAAAACACCCATACACATATAATTAAGAAACCTTAAGGTGATGT	2082
Db	121566	ACATGCAAGCAAGC--AAACCCCATATACACATATAATTAAGAAACCTTAAGGTGATGT	121624
QY	2083	GTGTGTAAACATTTGTCTTACACATGCTGATTAAGAGACATGTACAAACGCAACAGTGAAG	2142
Db	121625	GTGTG-----TGTGCAATACACATGTGTCTGTAAAGACATGTGTGCAACACACACACACA--	121675
QY	2143	AGGATCTGGGGCTGGAGATGTGCTCAGCGTTAAGAGCACTGACTGCTCTTCCGAAG	2202
Db	121676	-----	121675
QY	2203	AAGTCTGTGATTCAAATCCTAGCAACCACTATGTGGCTGCACAAACCATCATATATAGAT	2262
Db	121676	-----CACACACACACACA-----	121691
QY	2263	CTGACACCCCTCTTGTGTGATCTGTAAAGACAGCTGACAGAGCTACAGTGTACTTAAATATA	2322

Db	121692	-----	121691
Oy	2323	CTAATAAATAATCTTTTAAAAAATGAGAGCATGTGAGACACCTCAAAAGAGAT	2382
Db	121692	-----CACACACTGGGAAGAGGCTCAGACACTTCAMAAATAGAT	121730
Oy	2383	TATGAGCAGTgACTCACGGGTGATTATCTATCTGTGAGTCTTTTCCGCTGGCTTG	2442
Db	121731	TATGGGCACTGACTCATGGGTGACTATCTATCTGTGAGTCTTTTCCGCTGGCTTG	121790
Oy	2443	CACCTGGGTGGACAGACGCCCTTTTCATTACAAAGAACGGGTGCTACATTATTCTCAA	2502
Db	121791	TAACTGGGTGGACAGGACCTCTTTCATTACAAAGAATGAGTGTATATTATTCTCAA	121850
Oy	2503	CAAAACACACCTGTGAGTATGTATCTCTCTGCTACTATGAGACAGCGCACGCGCGC	2562
Db	121851	CAAAACACACCTGTGAGTATGTATCTCTCTGCTACTATGAGCGCTGTGGGCGC---	121907
Oy	2563	GCGCACACACACACACACACACACACACACACACACACATTCAGTCTCCA	2622
Db	121908	-----CACACACACTTTCAGTCTCCA	121930
Oy	2623	GAGCTCTTGGGAAGGTCAAGAAAGAGCTGCCCTCAAAACAGATCTTACTTTCCCTCT	2682
Db	121931	GAGCTTTTGGGAAGGTCAAGGAAGAGGCGCCCTCAAAACCTCTTCAATCTTGGCTCT	121990
Oy	2683	AAAGGAGACACAGATTCCAGAGGTGGCAGAAATCTACAGGGGGCAGAGGCAGGGAGG	2742
Db	121991	AA--GGAACCAAGATTCTTAAGTGGCAAAAGTTCTACAGGGGGCCAGAGGCAGGGAGG	122048
Oy	2743	AAGCAGGCCATGTGTTCCAGAGACTACAGACAGAGGGCAGCAGAGGACATCCAGGTCC	2802
Db	122049	AAGCAGGACACAGTTTCCAGAGACTCTACAGCAGAGGGCAGCAGAGCAGATCCAGGTCC	122108
Oy	2803	AGGGCAGGAGAGGTGGAGGCGCTTGTTCGAGAGA-----GAAGGACGGGGGAGAAACAGGG	2857
Db	122109	CGGGCAGGAGAGAGAGGCGCTTGTCTCTGTGGAGAGAAAGAGGTTCCGGGAGAGCAGGG	122168
Oy	2858	TTCAAAAGCA--CAGGTTATGTGGCAGACTATAAAGTGGAGGTGCTGTGCTCACTCAGAAAG	2916
Db	122169	TTCAAAAGCATTTGTGTTATGTGGCAACATCAATAAAGGGAGTTGTGG-----CTCAGAAAG	122224
Oy	2917	GAGGAAGAAAGGAAAGGCCCTTTGTGCCACTGAGCGAGAGGTCAATGCTAGTAGAGAGAT	2976
Db	122225	GAGGAGGACTGGAAAGGTCTTTGTGCCACTGAGAGGAGGTCAATGCTACTAGAAAGAT	122284
Oy	2977	CTGCAGGGGTCCAGAGACGCCCACTGTCTGTCCCAAGGAAACCCCAAGTGTGAACCTGTG	3036
Db	122285	CTGAAGAGGTGCCAGGGGCCCACTGT-----CCCAAGGAAACCCCAAGTGTGAACCTGTG	122340
Oy	3037	GCTTGTGGTGTGATGATTCAGCTACAAAGACCCAGAGAGTCTCTACTCCATCCATCCAGT	3096
Db	122341	GCTTGTGGTGTGATGATTCAGCTGCAAGGCCCAAGAGTCTCTGTCTCTCTCTCTCTCTCTCT	122400
Oy	3097	GCCCCCTGCCCCGCCACA-----CCCCACCCCGCACTCCCGTGC	3136
Db	122401	NN	122460
Oy	3137	CACATTCTTAGGGCTGAG-----	3156
Db	122461	TACTTTTCTTAGGGCAGAGAGGTGTGTCTCTCAGATTCCCCCATCTTTGGCGGGAAGAGAA	122520
Oy	3157	-----GTGGCCAGCCCTGTGTGGGGGTTCCTTACTCTGAGTGTAGAGCCCAAGT	3203
Db	122521	TGTGTCCGACACTGTGGCTCTCCCTGTGGGAGATGTCTACTCTCAGGAGAAAGTCTCAGGT	122580
Oy	3204	CCTAGCGGAGTGCACCCCATCTCTGAAAGTGCAGAGGCCAAAGGGGGGACACAGCGCAG	3263
Db	122581	CCTAGCGGAGTGCACCCCATCTCTGAAAGTGCAGAGGCC--AGGGCGGAGACACAGCGCAG	122639
Oy	3264	CTCAGGCTGACAGCTGTGTCTGGGCTCTAGGTTCCAGAGGACCTGTGGCACTTACTTCCC	3323

Db	122640	CTCTGGCTGTCAAGCTGTTGCTGGGCTCTAGGCTCCACAGGACCTGGGCACTTCTCC	122699
Qy	3324	CACCCCCCATTCATCTCTCTGGGGCCCTATCTTCCTTATATGATGGAAGATTCCT	3383
Db	122700	CACCCCTCATCATCTCTCTCTGGGGCCCTATCTTCCTTATATGATGGAAGATTCCT	122759
Qy	3384	GGGGGGGGGGGCTGGTGTGTAGAGACAAGTCGTTCCGTCTCCGACGACCTTGCAC	3443
Db	122760	GGGGGTGGGGG-----GGTGAAGACAAGG---TCGGTCTCCTGCAGCAGCTTCCAC	122810
Qy	3444	AACCTTCCTAAGATCTCCAGGTGGGGGCGGCTCTTCCACAGATGAAGCAATTGGGTG	3503
Db	122811	AACCTTCCTAAGATCTCCAGGTGGGGGCGGCTCTTCCACAGATGAAGCAATTGGGTG	122870
Qy	3504	GGGACACATGATGACACAGGTGTGTGAGGGGACAGGTCCTGCTCTCTCTGGCAGC	3563
Db	122871	GGGACACATGATGACACAGGTGTGTGAGGGGACAGGTCCTGCTCTCTCTGGCAGT	122930
Qy	3564	CTGTGCTTCTGTAGCACTTGATTAAGTTGGGGGTAGGTTAAGTGTCTGTGAACATC	3623
Db	122931	CTGGAC-----AAGTTAGTGAAGAGGGGTAGGTTAAGTTCTCGGAATC	122976
Qy	3624	TCAAAGAAGCAGAAGCCAGCGGCGTCTGGGCTTCAATGAAGGAATTC--ACAA	3681
Db	122977	TCAAAGAAGCAGAAGCCTAGCAGCTGTCTGGGCTTAAAGAAGGAATTCCTTCCCT	123036
Qy	3682	CCCCCTTCTGTAGTCACTCTCGCTCATGTGTATGATTCCTCGGAGCACAAGGTGC	3741
Db	123037	GTAACCCACCTTCACCTCAAGTTGGGTCCATCTGTGCAGATTCCTCGGAGCTGA--GTGC	123095
Qy	3742	TCCTGGGACTCAGATTTCTACAATTAAAA--TCAGACAGTCTGAG-----A	3787
Db	123096	TCCTGGGACTCAGATTTTACAAATTAATCTCAGGACAGTCTGTGAAGAAGCTGAATTA	123155
Qy	3788	CTTGACATCCGCGCTGATTTACTACTCTCTCGGCGCTCATTTCTGTCTATGTC	3847
Db	123156	GTTGGTCTCTCTTGGCAACATTTCTTCTCTCTCTGGAAGTTCTTCTTGTATATGTC	123215
Qy	3848	TTACACATCTGAATATG-----TTCTCTGTGTACACATTCCTCGACTCTCGGAGGT	3903
Db	123216	TTGCACCTCTGAATATGTTTCTTCTTCTGTGTACACATTCCTCGACTCTTGAAGGT	123275
Qy	3904	CGTATCTTGGGAC---ATGTATCTGGGATGTAACTGCAGCACCAGA--GAGAGG	3958
Db	123276	TGGGCTCTTGGGCCACCGATGATCTGGGATGTAACTGCAGTGCACGAGGAGGGAAG	123335
Qy	3959	GGAGAGTCAGAG-----CTGTGTCTTAGGCCCTATTAGGCTCGAGATACCCCTTCT	4014
Db	123336	GCAGAGGTGGAGGCTACAGGGTCTTAGGCCCTCTCTGGCGGTATCATCTCTGCCCT	123395
Qy	4015	AGAAATGGCCCCCTCATTTTTCGTTTACCATATCTATTTTATACAGATGGGAGTGA	4074
Db	123396	AGAAATGGCCCCCTCTGCTTTCAGTCTCCGTGTATTT--TATCAGAGCTGACAGTGA	123453
Qy	4075	AAGCCAAACCTGCCCCAGAGTTTGGGACTCAGTCAGACCAAGTTATCTGTACAGAATC	4134
Db	123454	AAGCCAAACCTGCCCCAGAGTTTGGGACTCAGTCATCAACCAAGTCATATGTCAGAGTTC	123513
Qy	4135	CCCCGTGACTT--GAGGTGGGAGAAATGCTCTGAGGGGCTTCCAGCTCTTGGTTAGCA	4193
Db	123514	TTTTTGTACTTGGAGGTGGAGATATCTCCTCCAGGGGCTCCAGAGCTGTGGTTAGTA	123573
Qy	4194	GGAGGATATCTTTTATAG---GCATGACCTAGTCTATGGTGTACTACATTCCTGTGC	4250
Db	123574	GGAGGATATCTTTTATAGGAGCCTGACCTAACCCATGGTGTATCCACACCTCTGTGC	123633
Qy	4251	CAGTTTAAAGCTGGAATTAACCCAGCGGACGCCGCAAGATTCCTACAGTTGTACCC	4310
Db	123634	CAGTTTAAAGCTGGAATTAACCCGCTGGACACCCAGAGATTCCTTAATGTTGGACATCC	123693
Qy	4311	AAGAACAACAAGCATGATATGCAAGATATAGTAGCTGGGAGAGAAGAATTTAA--A	4369
Db	123694	AAGAACAACAAGCATGATATGCAAGATATAGTAGCTGGGAGAGAAGAATTCCTTAAGT	123753

OY		4370	CCCCCCAAAGGCCACAGGCTTCGGTTCCCTTACTTGATGCACATGCGCACTATGAGTGTAG--	4427
Db		123754	CTAAGGCCCATAGTGCCTCCTCCGTTCCCTTAGTTCGAATGTCACTGTGTAAAGCCAGTT	123813
OY		4428	-----CTACTATTGGCGCTGAGATTGGTGTAGCTAACAGCACTAGATGTA- TT	4480
Db		123814	GGAATTCCTACTATGGGCGCTGAACTGTGATGCTACAGGTGTGAGTGTGACTTCATGTGTGT	123873
OY		4481	AGTGTGTATAATCTGAG-CACTTGGAGGCTGGAAGCAGAGAGAATTCGTATATGTTTTAGG	4539
Db		123874	AAGTGTATAAATCTGAGCCACTGTGAGAGGCCAGCATG----TTTGGTATGTTTGAAG	123928
OY		4540	CCAGCGTGAGCTATGAGGCGAGACCTTGTCTTAAACAATAAAGAAAATGAAGCCCAGCGAGTGG	4599
Db		123929	CCAACCCTGAGCTATGAGGGAAGCGCTTGTCTTAAAAAAGAAAGAAAGCAAACCCT--	123986
OY		4600	TGGCACACGCGCTTTAATCCAGCACCTTGGGAGGCAAMACAGGCAATTTCTGAGTTCAA	4659
Db		123987	-----	123986
OY		4660	GGCCACGCCCTGCTATAGAGTGAAGTTCCAGAGACGCCAGGCGTACACAGAAACCCCTGT	4719
Db		123987	-----	123986
OY		4720	TTTTGAAAAACAGAAAAACMAAACMAAACMAAACMAAACCAACCAACCAACCAAA	4779
Db		123987	-----ACCTTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCAAAACCAA	124020
OY		4780	ACCTTCATCTCATCTCTCTAGGCTGTGTCTGTCTAGGTGTAGAGTTGGGACCTTC	4839
Db		124021	ACCAACAAAATTCATCTGTCTAGGTTGTGTCTGTCTAGGTGTAGAGTTGGGACCTTC	124080
OY		4840	AGACTTATATATAAATPAGSCCTTTTATATACGTGTCAGAGAGAGAGAAAGCTTACGTCTG	4899
Db		124081	AGACTTATATAAATPAGSCCTTTTATATACGTGTCAGAGAGAGAGAAAGCTTACGTCTG	124140
OY		4900	GGACACAGTGGGACCCCTGAGAAAGTACTCTCTCCAGCCCAAAATTTGGGAGGCTTC	4959
Db		124141	GGACACAGTGAAGTCCCTGAGAAAGACCTCTCCAGCAGCTAAGAAATTTGGGAGGCTTC	124200
OY		4960	CTGGGAGGAAGTGTGC-----	4975
Db		124201	CTGGGAGGAAGTTCCTCCGATGTGAGTCTGGAAGAAATGAAGTGGCAACCAAGACAGGA	124260
OY		4976	---CCGATCAGACTACTGTTCTAGAAAGCAGAAAGAGAGGTTGGAAGATGT-----	5024
Db		124261	GCAGCCTGTAGAACTACTGTTCTAGAAAGCAGAAAGAGAGGTTGGAAGATGTGCTGAGG	124320
OY		5025	-----TGTGTGACAGACAGTTGGCAACGAAGGACAG-----AGGGGAGGCGATCC	5070
Db		124321	GACAGAGCGATGTGTGACAGAGAGTTGGGACAGGGGCACTTGGGACAGGGGGAAACGCTGC	124380
OY		5071	AAGATTCTGAACATGATGACTGTTGGTTCTGTGGGTGAAAGAGTGTCCCCAGGAGATA	5130
Db		124381	AAGATTCTGTACACATAGCTGGCTTTGGTCTCCGGGTGACAAAGTGTCTCCACAGG---	124438
OY		5131	GGGCTGTGAAAAGGGGACAGGGGGGTGAGCCAAATGATTCAGTTGAGSACACATCCAGC	5190
Db		124439	-----ATGGGGACCCAGGGGGGTGAGTCACTGTGATTCMAAGTTGAGGAACATATCTACC	124488
OY		5191	CCAGGGCTCTGTGCGCAAGCTTAAAGAAATGAGAGCCCTTAACCCCTCGTGAAGTTTAG	5250
Db		124489	CCAGGGCTCTGTGCGCAAGCTTGAAGAAATGAGAG--CTTCAACCCCTCTTGAAGTTTAG	124547
OY		5251	GGAGACAGAGAGAGTGTAGGAGATCTCTTAGGTGAAAGAGAGAGTATCT-GCTGTACCA	5309
Db		124548	GGAGACGTGGGGAGCTGTAGGATGTCTTGTGAGTGTAGGGAGAGAGTGTCTCCCTGTGACCG	124607
OY		5310	ACATGTGCTAGAGACGAAGCA-----GTTGACACACTTACCCCTCAG	5351
Db		124608	ACATGTGCTGAGAGCAAGCACTTGTAGATATCCCAAGGGGTGTGGCCCGGCTCCTCTCAG	124667

QY	5352	AAACAGCCATCCCCCTTGGCTAAGAGAGCGTGGGCCCTTCTGTGTAAAGATCTTAC	5411
Db	124668	ACCCAGCCATCCACCTTGGCCCTAAGAGAGCGTGGGCCCTTCTGTGTAAAGATCTAC	124727
QY	5412	TTTCTTCAGAGAGAGCGCAGCAAGCTTGTGCCCTCCCTGGTGGCATTAACACCCCT	5471
Db	124728	-TTTCTTCAGAGAGAGCGCAGCAAGCTTGT-CCGCCCCCAACAGTCAATAAACATCCCT	124785
QY	5472	GTGTGTAACTTAATGTTTTATTTACTGTCAATGTGCTCCAGAGCAAGTCCCATCTGGAGAC	5531
Db	124786	GTGTGTAACTTAATGTTTTATTTACTGTCAATGTGCTCCAGAGCAAGTCCCATCTGGAGACT	124845
QY	5532	TCTGCTCCTAACTCAACAAAGTATGGCCCATCTCTCAACCCAGAGAGCTGCAGAGAGA	5591
Db	124846	TCCGTCTTGTGCTCAACAAAGTATGGCCCATCTCTCAACCCAG-AGTGTGCAAAAGAGA	124904
QY	5592	GCCTTAGAGAAAGGGTAACAGTAAACAAATATGGCCAGATTAATAACAAACATACTTCT	5651
Db	124905	GCCTTAGAGAAAGGGTAACAGTAAACAAACGCGCAAGATTAACAGAAATACTGTCTCT	124964
QY	5652	TTGTATCCCAATTTGGTTTTGCTGAACACAGAGGGGGTGTGTAGTATGTGTGTGTG	5711
Db	124965	TTGTATCCCAAG-----TTGCTGAACACAGA-----	124989
QY	5712	TGT	5771
Db	124990	-----GTGATGTAGTGTATGT	125020
QY	5772	TTTCAATGCAAAAGATATCTGATATTTGGGCCCATGTCCCAACAGGGTATTTGGGAGAGTC	5831
Db	125021	CTTCAATGCAAAAGATATCTGATATTTGGGGCCCATGTCCCAATAGGGGTATTCAGGTAAAGTC	125080
QY	5832	-----AGGCTTCTGCAACACAGTAAAGCTGCCCAAGATAGGATTTGGGGCTGAATC	5882
Db	125081	AGGCTTCTGCAAGCTTCTGCAACACAGCAAGCTGCCCAAGATAGATTTGGTGGCTGTAGATC	125140
QY	5883	ACCAAGGGGACGGTATCAGAGTGGAGCAAAATCAACAAGATAAAGCAACCCCTGTGGGGC	5942
Db	125141	ACCAAGGGGACGGTATCAGAGTGGAGCAAA- -CACAAATAAAGCCACCCGTGCAGGGC	125197
QY	5943	TCAGAAGAGGGATTTACAAGAGGTAAAGGCCAAGCCATTTATTCACAAAGATAGTCTC	6002
Db	125198	TCAGA- - -GAGTTTACAAGAGGTAAAGGCCAAGCCATTTATTCACAAAGATAGTCTC	125252
QY	6003	AAAAATCAAGTCAAGAGAGATTTAGCTGAGAGATGGGGCTGTCTGATGGGAGACACCTG	6062
Db	125253	AAAAATCAAGTCAAGAGAGATTTAGCTGAGAGATGGGGCTGTCTGATGGGATACCTG	125312
QY	6063	ACCTTGCACTTATTAAGTCACTAGGCCAAGAGACAGTCAACAGAGGGTACTGGGTCTTACT	6122
Db	125313	ACTTGTGCTTGTATGTACTAGGCCAAGAGACAGTCAACAGAGGGTACTGTACTTACT	125372
QY	6123	CAGCTTGGAGCAGGCACTGGGAATGGGTGACTTCATCTGATGTGAGAGAGGGCTGAGCA	6182
Db	125373	CAGCTTGGGCGAGCAATAGGAATGGG-----TGTGTGAGAGGGTTGTAGAA	125420
QY	6183	CCACCAAGTACAGTGTCC- -CTGTGTCTATGCAAGATATTCGTCGGGCGAGTTTTCAAA	6239
Db	125421	TGACCAAGTGCAGAGTGTCTCTGATGTGCTTATGTGTGGATATCCAGCAAGTTTTCAAA	125480
QY	6240	GGACTTAAGACTCATCTCGTGTGGAACAAAGTATCCAGAGCCCTAAGGCCCATTTTGTGTC	6299
Db	125481	GGACCCAGGTCTCATCTCGATGTGGAATTAAGTATTCAGAGCCCTAAGCCCTATTTGTGTC	125540
QY	6300	TAAATTAATCAGAACCCCTGGGGATGCAAGCTGTGACAGCAGAGAGCTTTTAAAAAGCT	6359
Db	125541	TAAATTAATCAGAACCTCTGGGGATGCAAGCTGTGAGCAGCAG- - -TTTAAAAAAGCT	125596
QY	6360	CCCAAGTGTCTGTGATGACG- -AGCGGAGCAAAACACACACTCAAGGTAAACAGAAAGA	6417
Db	125597	CCCAAGCGACTGTGATGACGGAATTCAGATTAACAAACACTCAAGGTAAACAGAAAGA	125656
QY	6418	GGCAAAAGCTAAGGAAAGCTGGGATAGGGAGCCCTTCTCAAGGCCAGTATGTGAGAGCTG	6477

[illegible]

D 126726 TGGGGCTGGAGGAGGAGGAGGATGCTTTCTTTGTATTATATCTATCTGATTAATCTAT 126785  
Q 7243 CATGCTTTCTTTGTATTAATAAGTAAGTA----- 7274  
D 126786 CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 126845  
Q 7275 ----- 7274  
D 126846 TCCATTCTTCTGTGACGCTTCTCCAGTGTAGGGTTACAGGCAATTCATCTACATC 126905  
Q 7275 CTCACAGATGATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7332  
D 126906 CAGACATTTTATTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 126965  
Q 7333 -----TACCTACCTACCTACCTATC----- 7352  
D 126966 CTATCTATCTAACTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 127025  
Q 7353 -----CATCCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7408  
D 127026 CTATCTATCTATCTAACTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 127085  
Q 7409 TGGGTCACTCTTACCCCTAGTGTGGGTTTACCAACACCCAGACATTTAT----- 7460  
D 127086 TATCTATCTAACTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 127145  
Q 7461 -----TTTTTTTGTATTTTATTAATCT 7486  
D 127146 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 127205  
Q 7487 AGGAGCTCAGGGGAGGAGCTGCTGTGCAATGCTAAACAGCTCTGCGACAGAGC 7546  
D 127206 AGAGATTAAAGGGGAGGAGCTTGGGCGCTGTGTATGTAACAAAGCTCTGTGACAGAGC 127265  
Q 7547 TGCAGCTCAGTCCCACTTTTGTTCAGGTGACTGTGAGACTGTGATCTGCAACGCG 7606  
D 127266 TCCAGGCCAGTCCCACTTTG-TCAGGTATCTGTGGGCTTGACATTTTCACAATGC 127324  
Q 7607 TATGAGCTCTCTCCACCTCCAGTTCACAGCTTCTGTGATCCAGTGGCGGCA 7666  
D 127325 TATGAGCTCTCTCCACCTCCAGTTCACAGCTTCTGTGATCCAGT-----GGGCAA 127380  
Q 7667 CTCTGAGCTCAGGAGGCGCTGTCCCTG-----TCTTCAGA 7703  
D 127381 CTCTGAGCTTACAGGAGCTGTCCCGTCCCGCCAGACCTAGGAACCACTCTCCACC 127440  
Q 7704 CCTACATATTTGCTGTGCAAGAGTTCATGTAATGAGATGCGTCTGTATTTCT 7763  
D 127441 CCTACAGATTTGCTGTGCAAGAGTTCATGTAATGAGATGCGTGTGTGCTGCTT 127500  
Q 7764 TATGCTGGCGCTTATCTTACACAGTTTGTGTGGGCCATGTGCTACTGCTATATCT 7823  
D 127501 TATGAGCTGCTCTTATCTACACGAAATTTGTG-----CCTACGCTACCTTACTC 127554  
Q 7824 TATCTATCATCTTATGAGCTTAAAGTGTCTGTGTGAGTAAACCACTTCTGT 7883  
D 127555 CATTTT-----ATCACTTATGAGCTTAAAGTGTCTGTGTGAGCAAAAACATTTCTGT 127611  
Q 7884 TTTCAATTAAGTGAATTTGTGGCCACCCCAACCCCTTTTATTTTATTTGAGACA 7943  
D 127612 TTTCAATTAAGTGAATTTGTGGCCACCCCAACCCCTTTTATTTTATTTT 127671  
Q 7944 AGGTCTTTCTGTGTAATCT-TGCAATCTGTGGTGTCTGTGAGCTACTGTGTAGAGGC 8002  
D 127672 TATTTTGTAGCTCTTCTATGAGTGTGGTGTCTGTGAGCTCAGCTGTAGAGGC 127731  
Q 8003 TGTGAGGCTGTCTCTCCACTTTTGACATCTCTGTGAACAGAGTGAAGCACTTCAAG 8062  
D 127732 TGTG-----GTTCCACTTTTGACATCTCTGTGAACAGAAATACCAATGAGCTCAAG 127783  
Q 8063 ACAATTTCTGT 8122  
D 127784 ACACTTTCT-----AGTTTTCTTATTTTGTGCGTGTGTGTGTGTGTGTGTGTGTGT 127832

Q 8123 TTTGTCTCTCAGTGTCTCAGT 8179  
D 127833 TTTGTCTCTCAGTGTCTCAGT 127892  
Q 8180 ATGTGCCATCTCTCAGT 8239  
D 127893 ATGTGCCATCTCTCAGT 127940  
Q 8240 TAGTGTGCCAAGTATCAGT 8299  
D 127941 TAGTGTGCCA-----GTGCAATCTAG 127963  
Q 8300 GCATTTAACTCAGT 8359  
D 127964 GGAATTAAGT 128018  
Q 8360 GCGCTGTCTGT 8419  
D 128019 CCGCTGT 128075  
Q 8420 GCTCACCAGTATCTCAGT 8479  
D 128076 GCTCACCAGTATCTCAGT 128135  
Q 8480 TCCAGAGT 8539  
D 128136 TCCAGAGT 128180  
Q 8540 AGATATAGAGGAGT 8584  
D 128181 AAGATATAGAGGAGT 128240  
Q 8585 GGT 8634  
D 128241 GGT 128300  
Q 8635 GGT 8686  
D 128301 GGT 128360  
Q 8687 TAAAGAGAGT 8746  
D 128361 TAAAGAGAGT 128420  
Q 8747 ACCTGT 8806  
D 128421 ACCTGT 128480  
Q 8807 ACTAAGGAGGAGT 8866  
D 128481 ACTAAGGAGGAGT 128540  
Q 8867 AGGTTAGGAGT 8925  
D 128541 AGGTTAGGAGT 128600  
Q 8926 CTCTGGGCTGT 8954  
D 128601 CTCTGGGCTGT 128660  
Q 8955 ACTGTCTCAGT 8995  
D 128661 ATGTGTCTCAGT 128701  
RESULT 5  
AC101018 149109 bp DNA linear HTG 21-AUG-2002  
LOCUS Rattus norvegicus clone RP32-328P7, \*\*\* SEQUENCING IN PROGRESS:\*\*\*  
DEFINITION 9 unordered pieces.  
AC101018  
VERSION AC101018.2 GI:22381320





QY 562 ACACCTGGCACTTTAAACACTTATTAACACTGCGCAGCGCTTCAGGTTTGAAGATCACT 621  
Db 97091 -----GTTTTTAATATCTGGCAGACCCCTCAGGTTGAAGATCACT 97134  
QY 622 TTCAACACAGAAAGAGTGCCTGCTGCTCAGCGTAGCGACACTGGCTCAGA 661  
Db 97135 TTGGAACACAGACAGAAAGTCTCGC-----CCTTGGTGTATGTGAACACTGGCTCAGA 97190  
QY 682 AGAGTATATTTAGTGAAGTACCTTCACATATCTTTGCACTTATCACTACAGCTGT 741  
Db 97191 ACAGGCGCTGTTAATGAAGTTGCTGCAACAACATCTTTGCACAGAGCCATACATG 97250  
QY 742 CAAATGCTACTCCCTAGTCCAGAGATGCGTGTACACTGCTTCTGCTTCCATCT 801  
Db 97251 CAACCGGCTAACTCCTTATCCACAGATGCGCTGTACACTTCTCTGTTCCATTA 97310  
QY 802 GGTGACATTTTGCACACAGAAATTAGAAATGTGGTATTTATTTGTGTGCTGAGAC 861  
Db 97311 GGTGACATTTTATCATTAACAGAAATTAGAAATGCTATTTATTTGTGTGCTGAGAC 97370  
QY 862 ACCATCCAGGCGCTTTTACATTTTACAGCAGATGTTTACTTACTGGGCTACTTCCAAC 921  
Db 97371 AACATCCAGGCGCTTTACACTTACAGCAGATATGTTTACT-ACTGATTAACCTCTAGC 97429  
QY 922 GGTTTAAACATTTTGTATATTTACTTATTTTGTGTGATGAGTGA--GCATGTAT 979  
Db 97430 CATTTAAGACAAATTTTTCATATTTACATTTTCTGCGCTGAGCTTGGAGTCTATAT 97489  
QY 980 ACATATGTATAGAGTATGATGATGCTGCTGCTACCTCAAAATCATTTG-----CAGA 1031  
Db 97490 ATGTATGATATATGATATGACGCGTGGGTGTGTATACCTCAAAATCAAGAGACATCAGA 97549  
QY 1032 TCCCCAGCAAGTGAATGACCGAGCGTTGATGTTATGTGGAAGCTGGAGCCAAAGC 1091  
Db 97550 TCCCCAGCAAGTGAATGACCGAGCGTTGATGTTATGTGGAAGCTGGAGCCAAAGC 97609  
QY 1092 TGGGCTCTGCAAGAGCAGCCAGTGGCTTAAACATGGAGACAGCTCTCTAGCCTAAG 1151  
Db 97610 TGGGCTCTGCAAGAGCAGCCAGTGGCTTAAACCTGGAGCGTCTCTCTAGCCTAAG 97669  
QY 1152 GTAATCTTATGTTTTTAAATAATATATTTCTCAGCCGGGTGTGTGGCAGACGCTTTA 1211  
Db 97670 ACAATTTTCAATTTCTTAAAAAATATATTTCTCTGGGGTTGGGATTTAGCTCATGTGATG 97729  
QY 1212 ATCCAGCACTTGAGAGCTGAGGTGAGAAATATATACACACAGCCAGCTGGGGTGCAG 1271  
Db 97730 AGTGTGCTGAGCAAGCCAGGCGCTGAGTTCAGTCCCAAGCTCCGAAAAAAGAAAA 97789  
QY 1272 AGCTTGGCCCTGTTTTTTTGTGTTTCTTATGTCACAGTGGTCTTACCTGCGGTAT 1331  
Db 97790 AAGAAAAAATATATATATTTCTGTGTGATCTGTGACACATGCTATATTTCCAGCAC 97849  
QY 1332 GT----- 1333  
Db 97850 TTGAGAGGGTGGGGTGTGGGATCATAAATTCAGTTGGGGTAAAGAGTGGCCCTGTT 97909  
QY 1334 -----CGGTGCAAGGGTGTCAAGATCCCTTGAAGCTGAGTTAAAGACAGTTGTATCAC 1387  
Db 97910 TTCCGCCCTTATATGCACTGCTGTCTTGGCTGTGTGTGAGGAATCGATCCCTG 97969  
QY 1388 GCGGCGCTTACAGATGCTGAAATTTGAACCAAGGCTCCCTAGAGAAAGACCCAGTGC 1447  
Db 97970 GAGCTGCTATGAGTGTCTGGAATTTGAACCCAGACCCCTGGAAGAGAGCCAAATGCTC 98029  
QY 1448 TTAACTTCTGAGCAGCCCTCCAAACCTCTTTTGAAGACTTTAACTTTTGTATATG 1507  
Db 98030 TTAACTTCTGAGCAGCCCTCCAAACCTTTAAATTTCTTAACTTTTGTGTGTG 98089  
QY 1508 TGGGAAGTGAAT-----GATCTTGCACTTACCAAGTGTGTG 1544  
Db 98090 GTGTGTGTGTGTGTATATGAAATTTGAACCTTAAAGGCTTGTACACACTAAGTGTGTG 98149  
QY 1545 CTGCGGTGTATGATCTGAGCGCCGTACCCACACAGCTAGTGTATACAGTTTAAAGGCAA 1604

Db 98150 CTGTGCTGAAGTATCTACTGAGCCACACCCAGCTGATGTGATATAGCTAATGGCAA 98209  
QY 1605 ACATTTAAACATGACATATAGTTGATATGATTTGAATA--TAGTCTGAGCTATTTGTTA 1662  
Db 98210 ACATTTAGCATATGACATATGATTTGGTATGAGTTTGAAGACCCAGTCTGTAGACTATTTGTTA 98269  
QY 1663 GCGTGAACCTTGTGCTGTCTTACAGTGTGCTGTGAGAAGATAGAAAAATGAAGACTTGAGT 1722  
Db 98270 GCGTGAACCTTGTGCTGTCTTACAGTGTGCTGTGAGAAGATAGAAAAATGAAGACTTGAGT 98323  
QY 1723 CTAGTCTGGAACCCACAGAGCGAGAGACCCACTCTGTAAGTGTGTCTGTAGCT 1782  
Db 98324 TGAATCCCGAAGCTCAGTGTGAGTGTGAGAACCCACTCTGTAAGTGTGTCTGTAGCT 98382  
QY 1783 TCATATCAACTTACATATATAGTTTCAATGATATATATATAGTAAATCTTTTAA 1842  
Db 98383 CTGCTATACAAAC-----GTAAGATTTATATGACATATGATATATAGTAAATCTTTTAA 98438  
QY 1843 GGTATATGTTGGAGGAGAGATGCTCAGCTTCCAGAGACACTTGTCTTTCAGAG 1902  
Db 98439 GGTATA--GTTGGGACAGAGAGATGGCTCAGCTGCCAAGAGCACTTGTCTTTCAGAG 98494  
QY 1903 GACCTGATTTCACTTCCAGAGACTCATATGTGTGCTCAGAGCCATCTGTATATCCAGTTG 1962  
Db 98495 GCGCTGATTTCACTTCCAGAGACTCATATGTGTGCTCAGAGCCATCTGTATATCCAGTTG 98553  
QY 1963 CAGAGGTTTCCACACCTCTTCTGGGCTCCAGAGACACACATATGATATGATACAGATAT 2022  
Db 98554 CAGAGGTTTCCACACCTCTCTTGGGCTCCAGAGACACACATATGATATGATATGATACAGATAT 98613  
QY 2023 ACATGAGGCAAAACACCATATACACATATATATATATGAGAACTTAAAGGTCATGT 2082  
Db 98614 ACAGCGAGGC--AAACCCCATATACATATATATATATATATGAGAACTTAAAGGTCATGT 98672  
QY 2083 GTTGTGAATCATGCTTCTTACATATGCTATTTAAACATGTTCAACAGGACACACAGAG 2142  
Db 98673 GTTGTG-----TGTGATATACATGCTCTGTAAACACATGTTCAACAGGACACACAGAG 98723  
QY 2143 AGGAGCTGTGGGCTGAGAGATGCTCAGCGGTTAAGACACTGCTCTTCCGAGG 2202  
Db 98724 ----- 98723  
QY 2203 AAGTCTGAGTTCAAAATCCTTAGCAACACATGCTGCTCAGACACATCATATATGAT 2262  
Db 98724 -----CACACACACACACAC----- 98739  
QY 2263 CTGACACCTCTTCTGTGTGATCTGAAGACAGCTGAGAGCTTACATGTATATATAT 2322  
Db 98740 ----- 98739  
QY 2323 CTAAATAAATATCTTTTAAATAAATGAGAGGATCTGAGACCTCAAAAGAGAT 2382  
Db 98740 -----CACACATGAGAGAGAGAGCTGAGACACTTCAAAATATGAT 98778  
QY 2383 TATGACAGTACTCAGCGGTGATTTATCTATCTGAGATTTTCTTTCGCTGCTG 2442  
Db 98779 TATGCGCACTGACTCATGTGGTCACTATCTATCTGAGATTTTCTTTCGCTGCTG 98838  
QY 2443 CAACGTGGTGGAGAGAGCGCCCTTTCATTTCAACAAAGGAGGCTTATATTTGTA 2502  
Db 98839 TAACTGGGTGGAGAGCACTCTCTTTCATTTCAACAAATGAGTGTATATTTTAA 98898  
QY 2503 CAAAGAGCACTGAGATATGTTTACTGTCTTGTGACTATGAGCAGCGGAGCGGCG 2562  
Db 98899 CAAAGAGCACTGAGATATGTTTACTGTCTTGTGACTATGAGCAGCGGAGCGGCGGCGG 98955  
QY 2563 GCGACACACACACACACACACACACACACACACACACATTCAGTCTCA 2622  
Db 98956 -----CACACACACATTCAGTCTCA 98978  
QY 2623 GAGCTTGTGGAAGTCAAGAGAGGCTGCGCTCAACAGAGATCTTATCTTCCCTCGC 2682



Db 98979 GAGCTTTTGGGAAGGTCAAGGGGAAGCTGCCCTCAACACCCTCTTCATCTTCCCTCCT 99038  
Oy 2683 AAGAGAGACACGATTTCCAAAGGTGGCAGAAATCTACAGGGGGGAGAGGAGGAGGGG 2742  
Db 99039 AA--GGAAACAAGATTTTAAAGTGGCAGAAATTTCTACAGGGGCCAGAGGAGGAGGGG 99096  
Oy 2743 AAGCAGGCGATGGTTTCCAGAGACCTACAGCAGAGGCGCAGCAAGGCAGATCCCAAGTCC 2802  
Db 99097 AAGCAGGACACAGTTTCCAGAACTACAGCAGAGGGGAGCAAGGCAGATGCCCAAGTCC 99156  
Oy 2803 AGGCGAGGAGGTGAGAGCCCTTGTTCGAGG----AGAAGCAGAGGCGCAGAAAGAGG 2857  
Db 99157 CGGGAGGGAGAGAGAGCCCTTGGCTCTCGGAGAGAGGAAGGTTCCGGCAGACAGAG 99216  
Oy 2858 TTCAAAGCA--CAGTTTATGGCAGCTCATAAAGTGGAGTGGGCTGACATCGAAGAA 2916  
Db 99217 TTCAAAGCATTTTGGTTTATGGCACTCATAAAGGGAGTTGTGG----CTCAGAAAG 99272  
Oy 2917 GAGGAAGAAGGAAAGGCCCTTGTGCCCACTGAGCGAGGCTCATCTGATGAGAGAT 2976  
Db 99273 GAGGAGAGACTGGAAAGTCTTGTGCCCACTGAGCGAGGCTCATCTGATGAGAGAT 99332  
Oy 2977 CTGAGGGGTGGCAGAGACCCCACTGTCTGTCCCAAGGGAACCCCAAGTGAAGTCTG 3036  
Db 99333 CTGAAGAGGTGCCAGGGGCCAC---CTGTCCAAAGGGAACCCCAAGTGAAGTCTG 99388  
Oy 3037 GCCTTGGGTGCTGATTCAGAGTACAGACCCAGAGAGTCTACTCATCCCATCCAGT 3096  
Db 99389 GCCTTGGGTGCTGATTCAGAGTGCAGAGGCCCAAGAGTCTGTGCTCATCCATCTCGT 99448  
Oy 3097 GC----- 3098  
Db 99449 CCCCANN 99508  
Oy 3099 ----- 3098  
Db 99509 NNN 99568  
Oy 3099 CCCCCTGGCCCCCAGACACCCCAACCCCGACCTCCGTGCATCTCTAGAGGTGAGG-- 3156  
Db 99569 CCCCCCACCACACCATCCCTCCCCCAACCCCGTACTTTCTAGAGGAGAGGT 99628  
Oy 3157 -----GTGGCCAG 3164  
Db 99629 GGTCTGTGAGTTCCTCCCATCTTGGCGGGGAAGGAAAGMATGTCGGGACCTGTGGCCTG 99688  
Oy 3155 CCCTGTGGG--GGTTGCTTACTCTGAGTGAAGCCCAAGTCTCTACCGGAAGTGCACCC 3222  
Db 99689 CCCTGTGGGAAAGTGTCTACTCTGAGGAGAGCTCAGTCTTGAAGGAGTGCACCC 99748  
Oy 3223 CATCCCTGAAGCTGAGAGCCAAAGGGGGGAGCAGCGGAG----CTCAGGCTGTGAGGC 3278  
Db 99749 CATCCCTGAAGCTGAGAGCCAGAGGCGGAGCAGCAGCCAGCAGTGTGAGTCAAGNC 99808  
Oy 3279 TGTGTGGGTCTGAGTTCCTCCAGGAGCTGGGACCTAATT--CCCCACCCCCCATCC 3336  
Db 99809 TGTGTGGGTCTGAGTTCCTCCAGGAGCTGGGACCTAANTTCCCNACCCTCATCC 99868  
Oy 3337 ATCTCTCTGGGGCCCTATCTTCCCTTATATG--TGAAGGAAGTTCCTGGGGGGGGGG 3394  
Db 99869 ATCTCTCTGGGGCCCTATCTTCCCTTATATGNTGAAAGTTCCTGGGGGGGGGG 99928  
Oy 3395 GTGGGTGAGAGCAAAAGTCTGTGGTCTCTGAGCAGAGCTTCCCAAACTTCTTAAG 3454  
Db 99929 -----GTGAGAGCAAAAG-----TGGTCTCTGAGCAGCTTCCCAAACTTCTTAAG 99979  
Oy 3455 ATCTCCAGAGTGTGGTCTCTTCCAGAGCAAGTGAAGCAATTGGTGGGAGACATGG 3514  
Db 99980 ATCTCCAGAGTGTGGTCTCTTCCAGAGCAAGTGAAGCAATGGGTGGGAGACATGG 100039  
Oy 3515 TGACACAGAGTGTGAGAGGGAGAGGCTCTTCTCTCTCTGAGCAGCTGTCTTCT 3574  
Db 10040 TGACACAGAGTGTGAGAGGGAGAGGCTCTTCTCTTCTTCTTCTGCACTGTGAC----- 100094

Oy 3575 GTAGCACCTGTGATTAAGTTTGGGGGTGAGTAAGTGTCTGAAACTCTGAAGAAGCA 3634  
Db 10095 -----AAGTTAGTAGAGGGGTGAGTAAGTATTCGGGAATCTCAAAAGCA 100145  
Oy 3635 AGAAGCAGCAGGCTGTCTTGGGCTTCAATGAAGAAATTC--ACAGACCCCTTCTCT 3692  
Db 100146 AGAAGCTAGACACTGTCTTGGGCTTAAATGAAGAAATTCCTTCCCTGTAACCCACT 100205  
Oy 3693 GTAAGTCACTTCTCATCTGTAGATTTCCCTGGGACCAAGTGGTCTCTGGGACTC 3752  
Db 100206 TCACCTCAGGTTGGGTCCATCTGTGACAGATTCCTGGGACTGA--GTGGCTCTGGGACTC 100264  
Oy 3753 AGATTTTCACAATTAATA--TCAGACAGTCTGAG-----ACTTGGCTCG 3798  
Db 100265 AGATTTTCACAATTAATAATCTAGAGACAGTCTCTGAGAAAGCTGAATTAAGTGTCTCT 100324  
Oy 3799 TGCTGTATTTACTCTCTCTGTGGCTGATTTCTGTGTCTTCAATGCTTACACATCTG 3858  
Db 100325 TGCAACAATTTCTTCTCTCTGTGAGGTTCAATTTCTGTATGCTTGCACATCTG 100384  
Oy 3859 AATAGG----TTTCTTGTGTACCATTCCTCCAGACATCTGGGAGGTCGTATCTTGG 3914  
Db 100385 AATAGGTTTCTTCTTGTGTACCATTCCTCCAGACATCTGTGAGAGTTGGCTCTTGG 100444  
Oy 3915 CAC----ATGTATCCTGGGATGTAAAGTGCAGCCACAGGA--GAGAGGGGAGAGTCAAG 3969  
Db 100445 CCCACAGATGATCTCTGGATGTAAAGTGCATCAGACAGAGGAGGAAAGCAAGTGTG 100504  
Oy 3970 AG----CTGTGTCTAGAGCCCTTATTAAGGCTGGAGATCACCCCTTCTTCTAGAATGGCC 4025  
Db 100505 AGGCTACAGGGTCTAGGCCCTCTGTGGCTGTGATCTCTGCTCCCTAGAAATGGCC 100564  
Oy 4026 CTTCATTTTGGGTATCATGATGATTTTATATCAGAGTGGGCGAGTGAACCAACT 4085  
Db 100565 CTCTCTTCTCAGTCTCGGTGTCTATTT--TATAGACTACAGTGAACCAAGCT 100622  
Oy 4086 GCCCAAGATTGGGACTACACAGCAAGGTTATCTCTGAGAAATCCCTCTACT 4145  
Db 100623 GCCCAAGATTGGGACTACACAGGTTATCTCTGAGAAATCCCTCTACT 100682  
Oy 4146 T--GAGTTGGGAATCTGCTCTGGGGCTTCAAGTCTTGGTTAGCAGAGGATTC 4204  
Db 100683 TGGAGTTGGGAGTATCTCCCTCAGGGCTCCAGGGCTGTGTTAGCAGAGGCAATCC 100742  
Oy 4205 TTTGTATAG--GCATGACCTAGTCTAAGTGTATCTATCTCTCTCTCAATTAAGC 4261  
Db 100743 TTTGTATAGGAGGCTGACCTAACCCATGAGTGTATCCACACTCTCTCAATTAAGC 100802  
Oy 4262 TGAAGTAAACCCACAGCGAGCGCAGGATCTCTACAGTGTACCCCAAGCAACAA 4321  
Db 100803 TGGACTAAACCCCTGGAGAGCAGCAGATTCCTTAATGTGAGTCCCAAGCAAGC 100862  
Oy 4322 GACAGTATATGCAAGGATAGTACGTGGGAGAAAGAACTTAA--ACCCCCCAAG 4380  
Db 100863 GACAGTATATGAGGAGCAGAGTACGTGGGAGAAAGAACTTAAAGCTTAAGCCCAT 100922  
Oy 4381 GCCCAGAGTTCCTGCTTCTTCAATGCAATGCAATGAGTGAAG-----CTACT 4432  
Db 100923 AGGTCTCTCTCTCTCTCTTCTTCAATGCAATGCAATGCAATGCAATGCAATGCAAT 100982  
Oy 4433 ATGGCTGTGAGTGTGAGTCAAGCATGAGTGA--TGTTCATGTGTAGTGTATTA 4491  
Db 100983 ATGGCTGTGAGTGTGAGTCAAGCATGAGTGTGAGTGTGAGTGTATTA 101042  
Oy 4492 TCTGAG--CACTTGGAGGCTGAAGCAGAGATCTTATGTTTGAGGCCAGCTGAGC 4550  
Db 101043 TCTGAGCAGCTTGAAGGCCAGCATG-----TTGTGTATGTTTGAGGCCAGCTGAGC 101097  
Oy 4551 TATNAGGAGACTTGTCTTAAAGAAAAATGAAGCCAGCAGTGTGGCAACAGCC 4610  
Db 101098 TATNAGGAGAGCTGTCTTAAAGAAAAAGAAAGCAACCACT----- 101144





FEATURES	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
Db	8574	TTTTTGTGTGGTTGGTTTCCTCGCATAGATCAGGACATCGAATAGCT	-----	8628			
Db	105170	TTCTTGGTTGGTTGGTTTCCTTGTATTAAGTCAAGTAATCTGAATAAGTAGCTTA	105229				
Db	8629	-----AGCCTGGGCTACATATACATCTGTCTCAAAAAGCCTATA	-----	GAGTAGG	8675		
Db	105230	AGGCCAGGCTGGGCTACATAGAGATCTTCTCTCAAAAAGCCTGTAAAGGGGTTGTGGGG	105289				
Db	8676	GAGTCGAGGCTTAAGAAAGAGCCTTAAAGCCGGCTGTATAGACACAGAGATAGCCTGAC	8735				
Db	105290	GAGACTGAAGCTAAACAAGAGCTTAAAGCTAAATGTGTACAGACAGAGATAGCCTGGGC	105349				
Db	8736	TATATAGCAAGACCTGTTTCAAAAACATNGAGGAGAGGGGTATGTTTAAAGCTGGGCT	8795				
Db	105350	TACAGAGCAAGACCTGTGTTCAAAACATGAGGAGGAGGGGTGTGTTTAAAGTGGGCT	105409				
Db	8796	GTTTAAACAGGCTAAAGGAGGCAATGTATAGACATTTGATTAAGAAAGATCATCATCAAA	8855				
Db	105410	ATGTAAAGCGGTACTAAGAGAGGCAATGTATAGTGTGTTGATTAAGAGAGATCATCATCAAA	105469				
Db	8856	GCCGGGTGGGACAGGTTAGAGGTTGACTTACAGTGTCTCAAGACCCCA-TAGAGAGGCACT	8914				
Db	105470	GCTGGGTGGCTAGAGGTTAGAGGCTGACATATAGTGTCTCAGAGATCTCCAGTAGAAGACCAAC	105529				
Db	8915	TTCCCTTCTCTCTCGGGCCCTCA-----GCTCG	8943				
Db	105530	TTCCCTTCTCTCTCGGGCCCTCAACCTGTGGCTCTACCCGACAGACTTGGGCACTCTCG	105589				
Db	8944	GCTGACAGGCACTGCTCTCACATGCTTCTCTAGGCTGCTCCACCATG	8995				
Db	105590	GCTCAATGGCACTTCTCTCACATGCTTCTCTCAGGCTCATCCACCATG	105641				
RESULT 6	BV067438/c	779 bp	DNA	linear	STS 31-MAY-2003		
DEFINITION	S212P60465FAL.T0 CZECHII/El Mus musculus STS genomic, sequence						
ACCESSION	BV067438						
VERSION	BV067438.1	GI:31183233					
KEYWORDS	STS.						
SOURCE	Mus musculus (house mouse)						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
REFERENCE	1 (bases 1 to 779)						
AUTHORS	Wade,C.M., Kulbokas,E.J., Iir, Kirby,A.W., Zody,M.C., Mullikin,J.C., Lander,E.S., Lindblad-Toh,K. and Daly,M.J.						
TITLE	The mosaic structure of variation in the laboratory mouse genome						
JOURNAL	Nature 420 (6915), 574-578 (2002)						
MEDLINE	22354684						
PUBMED	12466852						
COMMENT	Contact: Kerstin Lindblad-Toh Whitehead Institute for Biomedical Research, Center for Genome Research 320 Charles Street, Cambridge, MA 02141, USA Tel: 6172521477 Fax: 6172580903 Email: kersl@genome.wi.mit.edu Primer A: None Primer B: None STS size: 779 Protocol: WGS-disccovery: Paired-end low-coverage whole genome shotgun reads were generated from 129SL/Svint, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the WGSv3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated as STS and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.						
Location/Qualifiers							

Source	1.. 779
organism="Mus musculus"	
/mol_type="genomic DNA"	
/strain="CECHII/Ei"	
/db_xref="taxon:10090"	
/map="1 31-720 75128449-75127262"	
/clone_lib="CECHII/Ei"	
<1..>779	
STS	
BASE COUNT	162 a 230 c 153 g 234 t
ORIGIN	
Query Match	6.3%: Score 564.6; DB 11; Length 779;
Best Local Similarity	88.3%: Pred. No. 1.5e-153;
Matches 693; Conservative	0; Mismatches 9; Indels 83; Gaps 4;
QY	4500 CTTGGGAGGCTGAGACGAGAGATTCATATATGTTTGAGGCCAGCTGAGCTATAGACGC 4559
Db	779 CTTGGGAGGCTGAGACGAGAGATTCATATATGTTTGAGGCCAGCTGAGCTATAGACGC 720
QY	4560 AGACTTGTCTTTAAGAAAAA---ATGAAGCCGACAGCTGTTGGCACAGCCTTTAA 4615
Db	719 AGACTTGTCTTTAAGAAAAA---ATGAAGCCGACAGCTGTTGGCACAGCCTTTAA 660
QY	4616 TCCGAGACCTTGGGAGGAGAGAGAGAGATTTCTGAGTTCAAGGCCAGCTGGTCTAT 4675
Db	659 TCCGAGACCTTGGGAGGAGAGAGAGATTTCTGAGTTCAAGGCCAGCTGGTCTAT 600
QY	4676 AGAGTGAATTCAGAGACAGCCAGGAGCTACACAGAGAAACCCCTGTTTGAAGAACAGAA 4735
Db	599 AGAGTGAATTCAGAGACAGCCAGGAGCTACACAGAGAAACCCCTGTTTGAAGAACAGAA 540
QY	4736 AACAAACAAACAAACAAACAAACCCCAACCCCAACCTCTATCTCTAT 4795
Db	539 AACAAACAAACAAACAAACAAACCA---AAACCCCAACCCCAACCTCTATCTCTAT 486
QY	4796 CTCCTCTAGGCTGTGCTGCTCTAGGCTGAGAGTTTGGGACTTCAGACTTATATATAAT 4855
Db	485 CTCCTCTAGGCTGTGCTGCTCTAGGCTGAGAGTTTGGGACTTCAGACTTATATATAAT 426
QY	4856 AGGCTTTTATATCAC---TGCTCAGAGACAGAGAAAGTTTCAGTCTGGACAGAGTGG 4911
Db	425 AGGCTTTTATATCACTGTGTGTGTCAGAGACAGAAAGTTTCAGTCTGGACAGAGTGG 366
QY	4912 ACCCTGAGAAAGTACTCCTCCAGCCAGCCAAATTTCTGGAGAGCTTCTGGAGAGTG 4971
Db	365 ACCCTGAGAAAGTACTCCTCCAGCCAGCCAAATTTCTGGAGAGCTTCTGGAGAGTG 306
QY	4972 TGTCGCCA----- 4979
Db	305 TGTCGCCAATGAGTCTGGAGAAATATGAGAGACTGGAAAGCCAGGACAGACAGCT 246
QY	4980 -----TCAGACTACTGTTCTAAGAGGACAGAGAGAGGTTGGAAGAT 5022
Db	245 GGGGTAAAGGCACAGACTCAGACTACTGTTCTAAGAGGACAGAGAGGTTGGAAGAT 186
QY	5023 GTTGTGTGACAGACAGTTGGAAACAGAGAGAGAGGGGAGGATCCAGATTTCTGAAC 5082
Db	185 GTTGTGTGACAGAGAGTTGGAAACAGAGAGAGAGGGGAGGATCCAGATTTCTGAAC 126
QY	5083 ATGTAAGCTGACTTGTGTTCTCTGGGTGACAAAGTGTCCCGAGGATATAGGGCTGTAGAA 5142
Db	125 ATGTAAGCTGACTTGTGTTCTCTGGGTGACAAAGTGTCCCGAGGATATAGGGCTGTAGAA 66
QY	5143 GGGAGACAGGGGTGAGCAATGAGTTCAAGTTGAGGACACATCCAGCCAGGGTCTTGG 5202
Db	65 GGGAGACAGGGGTGAGCCAGTGAAGTTCAAGTTGAGGACACATCCAGCCAGGGTCTTGG 6
QY	5203 CTGGC 5207
Db	5 CTGGC 1

AC119514	LOCUS	AC119514	203605 bp	DNA	linear	HTG 19-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-407A8, *** SEQUENCING IN PROGRESS	AC119514				
ACCESSION	AC119514	AC119514				
VERSION	AC119514.5	AC119514.5				
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.					
SOURCE	Rattus norvegicus (Norway rat)					
ORANISM	Rattus norvegicus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
AUTHORS	1 (bases 1 to 203605)					
Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalibechi, Y., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswalto, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, J., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Day-carroll, L., De Anda, C., Dedetich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duvall, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, E., Geer, K., Gill, R., Grady, K., Guerra, M., Guetara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hennadene, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulak, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C., Kwis, C., Krift, C. L., Lebow, H., Levan, Z., Lewis, S., Lopez, J., Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Ma, J., Lohensuhewa, L., Louised, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmood, M., Malloy, K., Mangun, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Mlivosic, J., Mlivosic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Natir, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokelimeh, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K., Paternack, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L. L., Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M. A., Reigh, R., Rellly, B., Rellly, M., Ren, Y., Reuter, R., Richards, S., Riggs, F., Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scheier, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvatsbeyn, A., Slasson, I., Sltter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svalok, A., Tabac, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.						
TITLE	Direct Submission					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 203605)					
AUTHORS	Worley, K. C.					
TITLE	Direct Submission					
JOURNAL	Submitted (28-Apr-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
REFERENCE	3 (bases 1 to 203605)					
AUTHORS	Rat Genome Sequencing Consortium.					
TITLE	Direct Submission					

[illegible]



shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GYXO  
Center clone name: CH230-152K9

----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 216595 bases at least Q40  
Consensus quality: 218663 bases at least Q30  
Consensus quality: 220078 bases at least Q20  
Estimated insert size: 225049; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 234549: contig of 234549 bp in length.  
1 234549

## FEATURES

source

1. 234549

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

## misc\_feature

/clone="CH230-152K9"

/complement(232578..233002)

/note="clone boundary"

## misc\_feature

/clone="end:17"

/complement(232578..233002)

/note="clone boundary"

BASE COUNT 56096 a 55263 c 55409 g 54605 t 13176 others  
ORIGIN

## Query Match

Best Local Similarity 67.4%; Score 188.6; DB 2; Length 234549;  
Matches 359; Conservative 0; Mismatches 134; Indels 40; Gaps 5;

1846 ATATGTTGGAGGAGAGATGGCTCAGCTCCAGAGACATGCTGCTTGCAGAGGAC 1905  
1519 ATTTGGGGGCTGAGAGATGGTGTAGTGTAGATCATGCTGCTCTCCAGAGGAC 1578  
1906 CTAGATTCAGTCCCGAGCATATATGTTGGCTCAGAGACATGCTGTAATCCAGTCCAG 1965  
1579 GTGGGCTCACTTTCTAGTACCATATAGTGGCTCACACACCTCTGTACTCCAGTTCAT 1638  
1966 AGGTTTCCACACCTTCTGCTGCTCCAGAGGACACACACATCATATAGTACAGAGAC 2021  
1639 GGGATCCAGTACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1698  
2022 TACATCCAGCAAAACACCATACACATATATATATAGGAAATTAAGAGTGCAT 2080  
1699 TATTCAGAGTCAAAACATCATATATATATATATATATATATATATATATATATAT 1758  
2081 GTCTTGTATTAACATTTGCTTACATGCTGATGTAAGACATGTACACGACACATCTGA 2140  
1759 GTGTATGTGTGTATATACACACACTATATATATATATATATATATATATATATATAT 1818

2141 AGAG-----GGATCTGGGGCTGAGAGATGAGCTCAGCGGTTA 2177  
1819 GTAGCCCTGCTTTCTTATAGACTCATCTAGGGGGCTGGAGAGGTGCTCAGCGGTTA 1878  
2178 AGAGCAGTACGCTCTTCTCCAGAGAGAGTCTGAGTTAAATCTAGACACATGCT 2237  
1879 AGAGCAGTACGCTCTTCTCCAGAGAGAGTCTGAGTTAAATCTAGACACATGCT 1934  
2238 GGCTCAGACACATCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2297  
1935 GGCTCAGACACATCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1989  
2298 CAGAGCTACAGTACTAGTAT 2350  
1990 ---GGCTCAGACGCTACTTAT 2039

## RESULT 9

AC094169/c

LOCUS

DEFINITION

AC094169

AC094169

AC094169.7 GI:30467804

HTG: HTGS\_PHASE1: HTGS\_DRAFT: HTGS\_FULLTOP.

## KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 240435)

## REFERENCE

Muzny,D.,Marie, Metzger,M.,Lee, Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alspbrooks,S., Amla,A., Anguiano,D.,

Anylebsch,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Beahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Burch,P., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

## AUTHORS

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Ande,C., Dederich,D.,

Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

## REFERENCES

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Georgievski,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolyet,A.,

## REFERENCES

Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorenzowa,L., Louisedge,H., Lozdo,R.J., Lu,X., Ma,D.,

Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,

Mamhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

## REFERENCES

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Muntadas,M., Murphy,M., Nair,L.,

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

Nwackemehe,O., Okunou,G., Olarinpoogon,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,

Plopper,F., Polinder,A., Popovic,D., Prims,E., Pu,L., L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regler,M.A., Reigh,R.,

## REFERENCES

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodery,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,

Shetty,J., Shwartsbeyn,A., Sisson,I., Sitter,C.D., Smjs,D.,

Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J.,

Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,

Taylor,T., Thomas,N., Thomas,S., Tinney,A., Trejos,Z., Usmani,K.,







[illegible]

table.	----- Genome Center
	Center: Baylor College of Medicine
	Center code: BCM
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>
	Contact: hgsc-help@bcm.tmc.edu
	----- Project Information
	Center project name: GSX
	Center clone name: CH230-196N8
	----- Summary Statistics
	Assembly program: Phrap; version 0.990329
	Consensus quality: 215928 bases at least Q40
	Consensus quality: 218585 bases at least Q30
	Consensus quality: 220210 bases at least Q20
	Estimated insert size: 221226; sum-of-contigs estimation
	Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
	-----
	* NOTE: Estimated insert size may differ from sequence length
	* (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> ).
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 3 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	* 1 53758: contig of 53758 bp in length
	* 53759 53858: gap of unknown length
	* 53859 253122: contig of 199264 bp in length
	* 253123 253222: gap of unknown length
	* 253223 262322: contig of 9100 bp in length.
FEATURES	Location/Qualifiers
source	1..262322
	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-196N8"
	3152..4104
misc_feature	/note="Clone boundary
	clone_end:5p6
	site:
	end_sequence:BH320154"
misc_feature	53859..54886
	/note="wgs_contig"
misc_feature	91022..92334
	/note="wgs_contig"
misc_feature	complement(193497..194191)
	/note="Clone boundary
	clone_end:17
	site:
	end_sequence:BH320153"
misc_feature	251042..253122
	/note="wgs_end_extension
	clone_end:17"
BASE COUNT	56211 a 52467 c 54612 g 58218 t 40814 others
ORIGIN	
Query Match	2.1%; Score 188.4; DB 2; Length 262322;
Best Local Similarity	58.4%; Pzed. No.6,4e-43;
Matches 395; Conservative	0; Mismatches 261; Indels 20; Gaps 6;
1799	ATTAATAGTTACCATATATAATTAATTAATTCCTTTAAAGTATATGTTGGAGC 1858
178963	ATTTTATGTAGCCACAAGAACAAATATGATTCATTTGGCACAAATGGTTGGGAAGCTG 178904
1859	GAGAGATGGCTCAGCTCCAGAGACACTTGTGCTCTTGCAGAGAGACCTAGATTCAGTTC 1918
178903	GAGAGACACTTAAGACAGCTAAAGAGAGTGGCGCTCTTGACAGAGACCTGGTTCAAGTTC 178844
1919	CCAGAGCTCATATGGTGGCTGCACAGCCACTGTAATATCCAGTTCCAGAGAGGTTCCACACC 1978
178843	CCAGAGCCCAACAGGCTGACTCAACCACTTTTAACTCTAGTCCACGAGGAGCCCAAGACC 178784

```

OY 1979 CTCTTTCGCTCCACAGCAGCAGCATAGTACAGACATATCATGACGCAACCA 2038
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178783 CTCTTTCGCTCCACAGCAGCAGCATAGTACAGCAGGATATACAGCATCAGG-AAAAA 178725
OY 2039 CCCATTACACATTAATAATAAGGAAGCTTAAGAGGTGATGCTGTGTAACATTTGTG 2098
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178724 CTCTACATATGAGCACTTGAGTAACTATTTTATGTTATGTTTCTTCAACATATGCA 178665
OY 2099 CTTACACATGCTGATTAAGACA---TGTACACAGCAGCAGCT---GAGAGGATCTGG 2152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178664 TGCAGCCAGAGACCATGACATTAATCTGGGGGAGAGGAGACAGCTCAAAAAGAAATTTG 178605
OY 2153 GCGTGGAGAGATGGCTTACAGCGGTTAAGACAGCTGCTCTTCCAGGAAGAGTCTCTGA 2212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178604 GCGTGGAGAAATGGCTGAGTAAATAGCAGTGAAGCTCTCCCA---AGTCTCTGA 178549
OY 2213 GTTCAATTCGTAGAACCATGATGGCTGACAAACATCATATATGATGATGACACCT 2272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178548 GTTCAATTCGTAGAACCATGATGGCTGACAAACATCATATGATGATGATGATGCT 178489
OY 2273 CTCTTGTGATCTGAAGACAGCTGACAGCTACAGTCTACTTATGATATCTATATAATA 2332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178488 CTCTTGTGATCTGAAGAC-----AGCTACAGTGTACTATCTAATAATA 178438
OY 2333 AATCTTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATGAGCAGT 2392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178437 AATATCTTTTAAAAAAGAGAAATATATATGCTNNNNNNNNNNNNNNNNNNNNNN 178378
OY 2393 GACTCAGCGGTGATTAATCTATCTGAGATTTTTCCTTCCGCTTGGCTGCACTGGGTG 2452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178377 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178318
OY 2453 GACAGACGCCCCCTTT 2468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178317 TATGAACTCTATCTT 178302

RESULT 11
AC114361/c 238330 bp DNA 1linear HTG 22-SEP-2002
LOCUS 3 unordered pieces.
DEFINITION
AC114361
AC114361.4 GI:23195300
VERSION HTG: HTG_PHASE1: HTG_DRAFT: HTG_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
    Eukaryota; Metazoa; Chordata; Rodentia; Sclurognathi; Muridae; Murinae;
    Rattus.
    1 (bases 1 to 238330)
    Muzny,D.,Marle,M.,Metzker,M.,Lee,A.,Abramzon,S.,Adams,C.,Alder,J.,
    Allen,C.,Allen,H.,Alsbrooks,S.,Amin,A.,Anguiano,D.,
    Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Bacca,E.,Baden,H.,
    Baldwin,D.,Bandaranaike,D.,Barber,M.,Barnstead,M.,Benamed,F.,
    Bisswa,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,
    Bryant,N.,Buhay,C.,Burch,P.,Burrell,K.,Cavazos,I.,Ceasar,H.,Cenier,A.,
    Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
    Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
    Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,
    Delgado,O.,Denson,S.,Detamo,C.,Ding,Y.,Dinh,H.,Divya,K.,
    Draper,H.,Dugan-Rocha,S.,Dunn,A.,Dunlin,K.,Duvall,B.,Eaves,K.,
    Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
    Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Poster,P.,
    Fraser,C.M.,Gabriel,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,
    Gebregiorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
    Gunaratne,P.,Haaland,W.,Hamill,C.,Hamilton,C.,Hamilton,K.,
    Harvey,Y.,Havlak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
    Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
    Hollins,B.,Howells,S.,Huiy,S.,Hume,J.,Idlebird,D.,Jackson,A.,
    Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jollivet,A.,

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Ledow,H., Leyan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louisedge,H., Lozano,R.J., Lu,X., Ma,J.,
Manshewari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
McWhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nsoekeme,O., Okunolu,G., Olanunnsagoo,A., Pal,S., Parks,K.,
Pasternek,S., Paul,H., Perez,A., Perez,L., Plankoch,C.,
Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.,
Piazo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sotelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Swalek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G., and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 238330)
Direct Submission
Submitted (08-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 238330)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21744367.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRHV
Center Clone name: CH230-45C9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 218627 bases at least Q40
Consensus quality: 221500 bases at least Q30
Consensus quality: 223329 bases at least Q20
Estimated insert size: 234329; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```



*	11436	14325	contig of 2890 bp	in	length
*	14326	14425	gap of 100 bp		
*	14426	17305	contig of 2880 bp	in	length
*	17306	17405	gap of 100 bp		
*	17406	20166	contig of 2761 bp	in	length
*	20167	20266	gap of 100 bp		
*	20267	23474	contig of 3308 bp	in	length
*	23475	23475	gap of 100 bp		
*	23575	26403	contig of 2829 bp	in	length
*	26404	26503	gap of 100 bp		
*	26504	29899	contig of 3396 bp	in	length
*	29900	29999	gap of 100 bp		
*	30000	34929	contig of 4930 bp	in	length
*	34930	35029	gap of 100 bp		
*	35030	41463	contig of 6634 bp	in	length
*	41464	41563	gap of 100 bp		
*	41564	47890	contig of 6327 bp	in	length
*	47891	47990	gap of 100 bp		
*	47991	56092	contig of 8102 bp	in	length
*	56093	56192	gap of 100 bp		
*	56193	63063	contig of 6871 bp	in	length
*	63064	63163	gap of 100 bp		
*	63164	70819	contig of 7656 bp	in	length
*	70820	70919	gap of 100 bp		
*	70920	82482	contig of 11563 bp	in	length
*	82483	82582	gap of 100 bp		
*	82583	98645	contig of 16063 bp	in	length
*	98646	98745	gap of 100 bp		
*	98746	117782	contig of 15037 bp	in	length
*	117783	117882	gap of 100 bp		
*	117883	1194825	contig of 31543 bp	in	length

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="Rp11-36003"
/clone_1lib="RPCT-11 Human Male BAC"
1..8144
/note="assembly_fragment
clone_end:SP6
vector_side:left"
8245..9350
/note="assembly_fragment"
9451..11135
/note="assembly_fragment"
11436..14325
/note="assembly_fragment"
14426..17305
/note="assembly_fragment"
17406..20166
/note="assembly_fragment"
20267..23474
/note="assembly_fragment"
23575..26403
/note="assembly_fragment"
26504..29699
/note="assembly_fragment"
30000..34929
/note="assembly_fragment"
35030..41463
/note="assembly_fragment"
41564..47690
/note="assembly_fragment"
47991..56092
/note="assembly_fragment"
56193..63063
/note="assembly_fragment"
63164..70819
/note="assembly_fragment"
70920..82482

```

[illegible]

ACCESSION AC021016  
 VERSION AC021016.4 GI:11120952  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 191754)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 JOURNAL Toward a complete human genome sequence  
 MEDLINE Genome Res. 8 (11), 1097-1108 (1998)  
 PUBMED 99063792  
 REFERENCE 2 (bases 1 to 191754)  
 AUTHORS Cordes, M., Maupin, R., Hawkins, M. and Boyer, E.  
 JOURNAL The sequence of Homo sapiens BAC clone RP11-378A13  
 TITLE Unpublished  
 REFERENCE 3 (bases 1 to 191754)  
 AUTHORS Waterston, R.H.  
 JOURNAL Direct Submission  
 TITLE Submitted (12-JAN-2000) Genome Sequencing Center, Washington  
 JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 191754)  
 AUTHORS Waterston, R.  
 JOURNAL Direct Submission  
 TITLE Submitted (08-NOV-2000) Department of Genetics, Washington  
 JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 191754)  
 AUTHORS Waterston, R.  
 JOURNAL Direct Submission  
 TITLE Submitted (09-MAY-2001) Department of Genetics, Washington  
 JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Nov 8, 2000 this sequence version replaced gi:7630969.  
 -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: saplens@wustl.edu  
 -----  
 Summary Statistics  
 -----  
 Center project name: H\_NH0378A13  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McHertson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
 The RCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catalanese, J.J. and de Jong, P.J. (1996) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)  
 VECTOR: pBACE3.6  
 NEIGHBORING SEQUENCE INFORMATION:

FEATURES  
 Source

The clone sequenced to the left is RP11-146N10; the clone sequenced to the right is RP11-36003. Actual start of this clone is at base position 1 of RP11-378A13; actual end is at base position 191754 of RP11-378A13.

The sequence RP11-378A13 from 156426 to 156433 is derived from a single m13 subclone. Size of the region was confirmed by PCR from BAC DNA.

misc_feature	1..224	/note="similar to EST AA947005 (NID:g3110400) oq45e10.s1"
misc_feature	1..224	/note="similar to EST A1184556 (NID:g3735194) qd60a10.x1"
misc_feature	1..224	/note="similar to EST A1457236 (NID:g4310105) tk01d06.x1"
misc_feature	1..216	/note="similar to EST AA782751 (NID:g2842082) aj08c02.s1"
misc_feature	1..216	/note="similar to EST AM020425 (NID:g5873955) df09e11.y1"
misc_feature	2..223	/note="similar to EST AA175052 (NID:g1756173) ms82f02.r1"
misc_feature	2..223	/note="similar to EST AV079022 (NID:g5198850)"
misc_feature	2..223	/note="similar to EST AM538472 (NID:g7180898)"
misc_feature	2..49	/note="similar to EST AV330904 (NID:g6370956)"
misc_feature	7..224	/note="similar to EST A1454294 (NID:g4292719)"
misc_feature	7..224	/note="similar to EST AM433991 (NID:g6965298)"
misc_feature	12..223	/note="similar to EST AM059787 (NID:g6652109)"
misc_feature	20..223	/note="similar to EST A1709077 (NID:g4998853) as86d11.x1"
misc_feature	60..220	/note="similar to EST AA870364 (NID:g2965809) vq44e05.r1"
misc_feature	67..223	/note="similar to EST AV330904 (NID:g6370956)"
repeat_region	447..660	/rpt_family="Alu"
repeat_region	1182..1425	/rpt_family="L1"
repeat_region	1442..1583	/rpt_family="MERL_type"
repeat_region	1820..1928	/rpt_family="Alu"
misc_feature	1948..2347	/note="similar to EST AA224009 (NID:g1844621) zr12g06.r1"
repeat_region	2162..2293	/rpt_family="Alu"
repeat_region	2297..2587	/rpt_family="Alu"
repeat_region	2597..2804	/rpt_family="Alu"
misc_feature	2606..2923	/note="similar to EST AW351985 (NID:g6849698)"
misc_feature	2630..3135	/note="similar to EST AA234699 (NID:g1859192) zs39g05.r1"
misc_feature	2804..3242	/note="similar to EST A1916611 (NID:g5636466) wa28f08.x1"
misc_feature	2976..3245	/note="similar to EST A1401669 (NID:g4244756) th24e10.x1"
repeat_region	3097..3284	/rpt_family="MIR"

	misc_feature	3546..3781	/note="similar to EST AA102341 (NID:g1647297) zm26n05.r1"
	repeat_region	3848..4127	/rpt_family="Alu"
	repeat_region	4187..4469	/rpt_family="Alu"
	repeat_region	4474..4690	/rpt_family="Alu"
	misc_feature	4723..5241	/note="similar to EST AW444974 (NID:g6986736) "
	misc_feature	4931..5413	/note="similar to EST AA100813 (NID:g1647248) zm26n05.s1"
	repeat_region	6006..6137	/rpt_family="MIR"
	repeat_region	6243..6394	/rpt_family="MIR"
	misc_feature	6967..7092	/note="similar to EST AW461516 (NID:g7031684) "
	misc_feature	7935..8054	/note="similar to EST AW461516 (NID:g7031684) "
	repeat_region	8294..8433	/rpt_family="MER1_type"
	misc_feature	8553..8604	/note="similar to EST AI391439 (NID:g4217443) tE96e02.x1"
	misc_feature	8556..8749	/note="similar to EST AW461516 (NID:g7031684) "
	misc_feature	9362..9733	/note="similar to EST AI391439 (NID:g4217443) tE96e02.x1"
	misc_feature	9476..9723	/note="similar to EST AA595679 (NID:g411029) nk66f01.s1"
	misc_feature	9997..10265	/note="similar to EST AI634626 (NID:g4685956) tz30901.x1"
	misc_feature	10003..10482	/note="similar to EST AI828100 (NID:g5448771) wK31g06.x1"
	misc_feature	10003..10482	/note="similar to EST AI890574 (NID:g5595738) wm87e04.x1"
	misc_feature	10003..10482	/note="similar to EST AI978960 (NID:g5803990) wT68e06.x1"
	misc_feature	10003..10482	/note="similar to EST AL047914 (NID:g4728747) "
	misc_feature	10003..10482	/note="similar to EST AW027349 (NID:g5886105) wE73e09.x1"
	misc_feature	10003..10457	/note="similar to EST AI972200 (NID:g5769026) wT63n01.x1"
	misc_feature	10003..10421	/note="similar to EST W48859 (NID:g1336988) zc42c08.s1"
	misc_feature	10003..10330	/note="similar to EST AW074215 (NID:g6029213) xb09d12.x1"
	misc_feature	10003..10249	/note="similar to EST AA143563 (NID:g1712935) z032d10.s1"
	misc_feature	10003..10247	/note="similar to EST AI583619 (NID:g4569516) ts16d09.x1"
Query Match		2.1%;	Score 188; DB 9; Length 191754;
Best Local Similarity		67.4%;	Pred. No. 7.9e-43;
Matches 507;	Conservative	0;	Mismatches 190; Indels 55; Gaps 15;
OY	3328	AGACCAAGGCGGCGACACGCGACGCTGTCAAGCTGTGTCGTGGGCTTAAAGTT	3297
DB	164808	AAAGCCAAAGGCGGGGACACA--TCAGCTCTCATTTGCAAGCGTg-TGAGAGCGCTGTGGCT	164864
OY	3328	CCCAAGGACCTTG--GCACCTACTTTCGCCACCCGCCCAATCATTCCTCTGTGGGGCCCTAT	3355
DB	164865	CCCAAGGACCTTGCGACGACCTTCCTCCACCCCTCTCA-----CTCTTGTGGGCCCTAT	164919
OY	3356	CTTCCTTATATGTAAGAAGATTCTCTGGGGGGGGGGGTGTGTGTAGACAACAAAGTc	3415
DB	164920	CTTCCTTATATGTAAGAAGCATTCCTCGATGATCTCCAGGTGTGACGTGGC	165030
OY	3416	GTTGGTCTCTCGACGACCTTGGCACAACTTCTTAAGATCTCCAGGTGTGGCTGGC	3475
DB	164973	--TGCTCTTTCGACGACGCTTGCACAAATTCCTCGATGATCTCCAGGTGTGACGTGGC	165030

QY	3476	TCCTTCACACAGGTAAAGCAATTGGTGGGACACATGATCACCACAGGTGGTTGAAGG	3535
Db	165021	TCGCCAAGACAGGTAAAGCAACCTGGTGGGAACATGAGAGACCTTGGTGGTGGAGGG	1650
QY	3536	GACAGGGGTCCT-TCGTTCTCTCTGGACGCC- GAGCTTTCTGTGACACTTTGGATTAAGT	3593
Db	165091	GACGGGGTCCCTCTGCTTTCTTCGGGCCAATTGGGGCCCTCTGGAGCACCTTTGGTAGCGGC	1651
QY	3594	TTGGGGGTGAGGTAAAGTGCTCT- GAAACTCTGAAGAAGCACAGAGCC--AGCAGGCTG	3650
Db	165151	TCGAAGGGAGGAGCTTGACACTCTGGATAGCCTGAGAGTGCTGGAGACCACAGTGTGCTG	1652
QY	3651	TCCTTGGGCGCTTCATGAAGAAAGTTCACAGACCCCTTCTCTGAA-- GTACACTTGGCT	3708
Db	165211	CCCTGGGGCTCTCCAGGAGAGACTTTTGGAGCCCTCGCCCCAGACAAATGGCTCTCTCAGTC	16522
QY	3709	TCATCTGTGTAGATTCCCTGGGACCAAGTGGGCT----- CCTGGGACTCAGATT	3757
Db	165271	TCCTACCTGTGTGGTAGGCCCATTCCTACCTTTGGTGGTGCACCATCCCAATTGGCCCTGGAACT	1653
QY	3758	TCCTACATTTAAATCAGACAGCTCTCTAG----- ACTGGACTCCGTGCTGT	3805
Db	165331	TCCTAGAGTTAAACACAGACAGTCTGGGACAGACCGGAGCTAGTTGGTCTCCCTGACCTC	16533
QY	3806	ATTTAAC-- TACTCTCTCTGGCTGCTCATTTCTGTTCATGTCTTAACATCAGACTGAAAT	3862
Db	165391	ACTCCCGAGGCCCTCCACACAGCTCATCTCTTGGCCGTGGCTTCTTCACTGTGAAGT	16544
QY	3863	GG--- TTTCTTTGTGTACCACTTCCCTGACACTCTCTGGAGAGTCTGATCTTTGGACAT	3919
Db	165451	GTTTTTTTCTTTGTGTACCACTTCCCTGTGACACTCCGGGAGCCTCCGGGTCTGGCCCTCT	16555
QY	3920	GTATCTCGGAGTGTAACTGACGAGCCACACAGA	3951
Db	165511	GGCTCCTAGCATATGAGACACAGCCACACAGA	165542
RESULT 14			
LOCUS	G76462	568 bp	DNA linear STS 06-SEP-2002
DEFINITION	S208P6543FH7.T0 129S1/SvImJ Mus musculus STS genomic, sequence tagged site.		
ACCESSION	G76462		
VERSION	G76462.1	GI:22727218	
KEYWORDS	STS.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Wade,C.		
TITLE	Polymorphism Structure in the Mouse		
JOURNAL	Unpublished (2002)		
COMMENT			
	Contact: Kerstin Lindblad-Toh		
	Whitehead Institute for Biomedical Research, Center for Genome		

**WGS-discovery:** Paired-end low-coverage whole genome shotgun reads were generated from 125S1/5V1mJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the M5Gcv3 C57BL/6J assembly and SNPs detected was carried out by SSAHA-SNP. 225,000 reads were annotated

as STRs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

Contact: Kerstin Lindblad-Toh  
 Whitehead Institute for Biomedical Research, Center for Genome  
 Research  
 320 Charles Street, Cambridge, MA 02141, USA

**TITLE** Polymorphism Structure in the Mouse  
**JOURNAL** Unpublished (2002)

REFERENCE  
1 (bases 1 to 568)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REMARKS	DATE	SOURCE
		Mus musculus (house mouse)

VERSION G76462.1 GI:22727218

tagged site.  
076463  
ACCESSION

LOCUS	G76462	568 bp	DNA	linear	STS 06-SEP-2002
DEFINITION	S20826543FH7.T0.129S1/SyTmT Mus musculus STS genomic sequence				

## RESULT 14

DD 100311 06CICCIWOCIAIOWOCMAUCACUCCACACCA 100342

QY 3920 GTATCCTGGGATGTAAGCTGCAGCCACCAGGA 3951

Db 165451 GTTTTCCTTCTGTCACACTTCCTGACACTCTGGGACCTCCGGGTCCTGCCTCCT 1655

[illegible]

3806 ATTAC---TACTTCTCTGTGGCTGCTCATTTCTGTGTTCATGTCATTACACATCTGAAT 3862

Db 165331 TCTAGACTTAAACCCAGACAGTGTCTGGGCAGACCGGACTAGTTGGTCTCCCTGACCTC 1653

[illegible]

QY 3709 TCATCTGTAGATTCCCTGGGACCAAGGTGGT-----CCTGGACTCAGATT 3757

Db 165211 CCCTGGGCTCCAGGAGAGACTTTTGCAGGCCCTGCCCCAGACATGGCTCTCTGAGTC 1652

3651 TCTTTGGGCGCTTCAATGAAAGGAGGTTCCACAGAGCGCGCTTTTCCTGCTGCT 3708

3594 TTTGGGGTGGAGTAAAGTCTCTT-GAAACCTGGAAGAAGCAGAAAGCC--AGCAGGCTTG 3650

Db 165091 GACGGGICCTCTGCTTTCTCGGGCACTTGGGGCCCTCTGAGACACTTGGTACGGCC 1651

3536 GACAGGTCCT-TGCCTTCCTCTGGCAGCCT-CTGCTTTCCTGTACACACTTGCTATAAGT 3593

34/01C1C8ACACAGG1AAGGCATTTGGG1GGGGGACACATGG1GACACACAGG1GG1TGG6666 3535

-1.19e Page 36

FEATURES	source	Location/Qualifiers
		1..568
		/organism="Mus musculus"
		/mol_type="genomic DNA"
		/strain="129S1/SvImJ"
		/db_xref="taxon:10090"
		/map_xref="1 22-501 7012032-7012511"
		/clone_lib="129S1/SvImJ"
		<1..>568
STS		
BASE COUNT	176 a 122 c 111 g 159 t	
ORIGIN		
Query Match	2.1%; Score 186.6; DB 11; Length 568;	
Best Local Similarity	69.2%; Pred. No. 6.5e-43;	
Matches	351; Conservative 0; Mismatches 129; Indels 27; Gaps 6;	
QY	1847 TATGTGGGAGGAGAGATGGCTCAGCTTCACAGAGACACTGTCGTCTTCAGAGAGACC	1906
DB	79 TGTATTTGCTTGTGTTAAGACTCAGTGTCMAAGACTGCTGCTCTTCAGAGGAGCC	138
QY	1907 TAGATTTCAGTCCAGAGCATATTTGGGGCTCAGACAGCATCTGTAAATCCAGTCCGA	1966
DB	139 AGGGCTCAGTTCTCATTAAGCCAGAGATGGCTCACACTATTTGTAATCCAGTCCAGG	198
QY	1967 GGGTTCACACCTCTTCTGTGGCTCCACAGGACCCACATFACA-TAGTACAGACATACA	2025
DB	199 GCATCTGGTGCCCTCTCTCTGGGCTTCACAGGAATTTCTATFACAAGTACACAGATCA	258
QY	2026 TGCAGGCAAAACCCATFACACACATTAATTAATAGGAACCTTAAAGTGCAATGTT	2085
DB	259 TGC-----AAACACATCATACATTAATAATCAATTCATTAATTTAAATACATTAAT	313
QY	2086 GGTAAACATTTGTGTTACACATGCTGATTTGAAGACATGTACAGGCACACACTGAAGAG	2145
DB	314 GAT-AACAAAGCTGTCCAGAAATTTAGATAATATATATGCTTAAATTTTCTG-----	366
QY	2146 GATCTGGGGCTGGAGAGATGGCTCAGCGGTTAAGACACATGCTCTTCCGAAGAGAG	2205
DB	367 --GTAGGGGGCTGGAGTGTAGTCTCAGTGGAAGAGCACTACGCTCTTCC---GAAG	420
QY	2206 GTCTCGAGTCAAAATCCTAGACACACATGCTGCTCAACCAATCCATTAATGAGATCTG	2265
DB	421 GTCTGAGTTCAATCCAGACACACATGCTGCTCAACCAATCTATATGAGATCTG	480
QY	2266 ACACCTCTTCTGTGTCATCTGAAGACAGCTGCACAGCTACAGTACTTATGATATCTA	2325
DB	481 ACTCCCTCTTCTGTGCTGTGAAGAC-----AGCTACAGTACTCTACATATATA	532
QY	2326 ATAAATTAATCTTTTTTAAAAAAATG	2352
DB	533 AATTAATCTTTTTTTTTTTTAAATGTTG	559
RESULT 15		
AC100085	195636 bp	DNA
LOCUS	Mus musculus clone RP23-36K2, WORKING DRAFT SEQUENCE, 6 unordered	
DEFINITION		
ACCESSION	AC100085	
VERSION	AC100085.3	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 195636)	
JOURNAL	Birren,B., Nusbaum,C. and Lander,E.	
REFERENCE	Mus musculus, clone RP23-36K2	
AUTHORS	Unpublished	
	2 (bases 1 to 195636)	
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,	
	Anderson,S., Batra,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,	
	Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,	

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearthano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gedyas, S.,  
 Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kanat, A., Karatas, A., Kells, C., LaRoque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., MacDonald, P., Major, J., Margus, N., Mathew, G.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,  
 Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunhthang, P., Plenne, N., Pollard, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuppach, R.,  
 Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G.,  
 Zainoun, J., Zempek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 195636)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cook, A., Cooke, P., Corum, B., Dearthano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fato, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyne, S.,  
 Graham, L., Grand-Pierre, N., Hafer, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
 MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrum, J., Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zempek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 17, 2003 this sequence version replaced gi:22004247.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Project name: LI3938  
 Center project name: LI3938  
 Center clone name: 36\_K\_2  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1071: contig of 1071 bp in length  
 \* 1072 1171: gap of 100 bp  
 \* 1172 3399: contig of 2228 bp in length  
 \* 3400 3499: gap of 100 bp  
 \* 3500 18021: contig of 14522 bp in length  
 \* 18022 18121: gap of 100 bp





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:01 ; Search time 15870 Seconds  
(without alignments)  
13775.599 Million cell updates/sec

Title: US-09-877-935-1  
Perfect score: 8995  
Sequence: 1 gatcgtgtgcacaaagaca.....ctctagctctgcacacatg 8995

Scoring table: OLIGO NUC  
Gapex 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_png:\*  
27: em\_gss\_vrl:\*  
28: gb\_gssl:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	498	5.5	507	28	AZ354198 1M0093E23
2	322	3.6	336	28	AZ657211 1M0533G15
3	262	2.9	307	28	AZ354559 1M0093E23
4	242	2.7	291	28	AZ821333 2M0094K06

5	206	2.3	495	28	BH086621	RPCT-24-2
6	137	1.5	639	28	AZ821403	AZ821403 2M0094I08
7	81	0.9	565	28	AZ747682	RPCT-24-8
8	80	0.9	373	10	BH813805	BH813805
9	79	0.9	713	28	AZ902196	RPCT-24-1
10	76	0.8	671	10	BH662563	BH662563
11	75	0.8	1083	11	AK086147	AK086147 Mus muscu
12	75	0.8	438	14	CD538197	CD538197 B0201B07-
13	75	0.8	540	28	AZ427733	AZ427733 1M0209H24
14	74	0.8	674	29	CNS0212Y	AL200887 Tetraodon
15	71	0.8	779	13	BU705960	BU705960 UT-M-FRO-
16	70	0.8	321	12	BM219657	BM219657 C0928C02-
17	67	0.7	499	28	BZ134506	BZ134506 C8230-385
18	67	0.7	525	28	BH078848	BH078848 RPCT-24-2
19	67	0.7	676	27	DR4814T	AL980584 Danio rer
20	67	0.7	705	29	BX184553	BX184553 Danio rer
21	67	0.7	713	29	BX154096	BX154096 Danio rer
22	67	0.7	722	29	BX133701	BX133701 Danio rer
23	67	0.7	726	13	BX076218	BX076218 BX076218
24	66	0.7	138	29	BX217769	AA078152 7H14G12 C
25	66	0.7	167	9	AA078152	AA078152 7H14G12 C
26	66	0.7	214	14	CA8489703	CA8489703 omykrltb0
27	66	0.7	296	14	CA302665	CA302665 taal15b02.
28	66	0.7	302	10	BB347599	BB347599 BB347599
29	66	0.7	371	28	BH125739	BH125739 RPCT-24-3
30	66	0.7	377	29	BX239417	BX239417 Danio rer
31	66	0.7	408	9	AI032142	AI032142 os76d08.s
32	66	0.7	410	28	BH398554	BH398554 AG-ND-148
33	66	0.7	448	29	BX165815	BX165815 Danio rer
34	66	0.7	460	10	BE606136	BE606136 f123C06.Y
35	66	0.7	507	10	BF468206	BF468206 UT-M-CO-
36	66	0.7	567	29	DR5M24S	AL747210 Danio rer
37	66	0.7	573	29	DR4C10S	AL742982 Danio rer
38	66	0.7	574	28	AZ833285	AZ833285 2M0115G17
39	66	0.7	584	28	AZ412557	AZ412557 1M0186N01
40	66	0.7	586	29	BX184359	BX184359 Danio rer
41	66	0.7	594	29	BX165614	BX165614 Danio rer
42	66	0.7	612	12	BI839570	BI839570 f939H09.Y
43	66	0.7	620	29	BX152328	BX152328 Danio rer
44	66	0.7	625	29	U68740	U68740 U68740 R1ft
45	66	0.7	667	29	BX159688	BX159688 Danio rer

## ALIGNMENTS

RESULT 1  
AZ354198/c 507 bp DNA linear GSS 02-OCT-2000  
LOCUS 1M0093E23F Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
DEFINITION clone UGCGIM0093E23 F, genomic survey sequence.

ACCESSION AZ354198  
VERSION AZ354198.1 GI:10465402

KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 507)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Unpublished  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0093 row: E column: 23  
Seq primer: CGTTGTAAACGACGCGCCAGT  
Class: Plasmid ends  
High quality sequence stop: 507.

## FEATURES

source

Location/Qualifiers

1. 507  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0093E23"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g114732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

125 a 139 c 105 g 138 t

ORIGIN

Query Match

5.5%; Score 498; DB 28; Length 507;

Best Local Similarity 100.0%; Pred. No. 2.1e-164;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
8431 CATCCAGAGTGGTACGACGAGGAGACTCAGGATATCTGCTGCTCCACAGTGT 8490
|||||
507 CATCCAGAGTGGTACGACGAGGAGACTCAGGATATCTGCTGCTCCACAGTGT 448
|||||
8491 AGAATTACAGGCATATCACTGCTGGAAGATTTTAACTGAACTCTGAGATAGCA 8550
|||||
447 AGAATTACAGGCATATCACTGCTGGAAGATTTTAACTGAACTCTGAGATAGCA 388
|||||
8551 GGCACCTTACCAATGAGAGGTTCTTTTGGTTGGTTTCCCTGCTGAATGATCA 8610
|||||
387 GGCACCTTACCAATGAGAGGTTCTTTTGGTTGGTTTCCCTGCTGAATGATCA 328
|||||
8611 GGCAGTCTGAAGATGTAGCGCTGCTACATCACTTGTCTCAAAAAGCCTATAGAG 8670
|||||
327 GGCAGTCTGAAGATGTAGCGCTGCTACATCACTTGTCTCAAAAAGCCTATAGAG 268
|||||
8671 GTAGGAGGTGAGGCTAAAGAGAGCCCTTAAGCCGCTGTATAGCAGACAGATAGCC 8730
|||||
267 GTAGGAGGTGAGGCTAAAGAGAGCCCTTAAGCCGCTGTATAGCAGACAGATAGCC 208
|||||
8731 TGCACTATATAGCAAGACCTTTTCAAAAACATGAGAGGAGGATATGTTTAAGTCT 8790
|||||
207 TGCACTATATAGCAAGACCTTTTCAAAAACATGAGAGGAGGATATGTTTAAGTCT 148
|||||
8791 GGGCTGTGTACAGGACTAAGGAGCAATGTAGACATTTGACTAAGAAAGATCATCA 8850
|||||
147 GGGCTGTGTACAGGACTAAGGAGCAATGTAGACATTTGACTAAGAAAGATCATCA 88
|||||
8851 TCAAAGCCGGGTGGGAGGATAGAGTTGACTACAGTGTCAAGACCCCATAGGAAGC 8910
|||||
87 TCAAAGCCGGGTGGGAGGATAGAGTTGACTACAGTGTCAAGACCCCATAGGAAGC 28
|||||
```

QY 8911 CAGTTCCCTCTCTCTC 8928  
|||||  
Db 27 CAGTTCCCTCTCTCTC 10

## RESULT 2

A2657211/c

LOCUS

DEFINITION

1M0533G15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

VERSION

A2657211.1 GI:11794357

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Contact: Robert B. Weiss

University of Utah

Blm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0533 row: G column: 15

Seq primer: CGTTGTAAACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 336.

Location/Qualifiers

1. 336

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0533G15"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PWD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g114732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

103 a 88 c 66 g 79 t

ORIGIN

Query Match 3.6%; Score 322; DB 28; Length 336;  
Best Local Similarity 100.0%; Pred. No. 2e-102;

(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold<sup>®</sup> (Stratagene) cells and selected for ampicillin resistance.

	Query Match	2.9%	Score 262	DB 28	Length 307
	Best Local Similarity	100.0%	Pred. No. 2,7e-81		
	Matches 262; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1128	TGGGACCAAGCTCTCTAGGCCCTAAGGTAATCTTTAGTTTAAAAATATATATTCACGC	1187		

FEATURES	SOURCE	TITLE
1. location/Qualifiers	1. 307	Mouse whole genome scaffolding with paired end reads from 10kb and Wright, D., Weiss, R.

organism="Mus musculus"	Journal	Plasmid inserts
/mol_type="genomic DNA"		Unpublished
/strain="C57BL/6J"	COMMENT	Contact: Robert B. Weiss
/db_xref="taxon:10090"		University of Utah Genome Center
/clone="U06C1M0093E23"		University of Utah
/sex="Male"		Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"		84112, USA
/clone_lib="Mouse 10kb plasmid U06C1M library"		Tel: 801 585 5606
/note="Vector: PWD424v; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource"		Fax: 801 585 7177
		Email: ddunn@genetics.utah.edu
		Insert length: 10000 Std Error: 0.00

RESULT 4	
LOCUS	AZ821333/c
DEFINITION	AZ821333 291 bp DNA linear GSS 20-FEB-2001
ACCESSION	2M0094K06F Mouse 10kb plasmid U0CC1M library Mos musculus genomic
VERSION	clone U0CC2M0094K06 F, genomic survey sequence.
KEYWORDS	AZ821333
SOURCE	AZ821333.1 GI:12991241
ORGANISM	GSS.
TITLE	Mus musculus (house mouse)
REFERENCE	Mus musculus
ATTNORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 291)
	Dunn,D., Acyragi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UTM 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00



REFERENCE 1 (bases 1 to 639)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingy, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0094 row: 1 column: 08  
 Seq primer: CGTTGTAAACGACGGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 639.

FEATURES  
 source  
 1. 639  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0094108"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (g114732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 163 a 166 c 121 g 189 t  
 ORIGIN

Query Match 1.5%; Score 137; DB 28; Length 639;  
 Best Local Similarity 100.0%; Pred. No. 2e-37;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2694 CGATTCAGAGTGTGCAAGATCTACAGGGGCGAGGCGGAGGCGAAGCGGCAT 2753  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 179 CGATTCAGAGTGTGCGAAGATCTACAGGGGCGAGGCGGAGGCGAAGCGGCAT 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2754 GGTTCACAGACCTACAGCAGAGGCGAGCAGATCCCGAGTCCAGGCGAGGAG 2813  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 119 GGTTCACAGACCTACAGCAGAGGCGAGCAGATCCCGAGTCCAGGCGAGGAG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2814 GTGAGGCGCTGTGTCC 2830  
 ||||||||||||||||||||  
 Db 59 GTGAGGCGCTGTGTCC 43  
 ||||||||||||||||||||

RESULT 7  
 A2747682 565 bp DNA linear GSS 25-JAN-2001  
 LOCUS A2747682

DEFINITION RPCI-24-80N12.TJB RPCI-24 Mus musculus genomic clone RPCI-24-80N12,  
 genomic survey sequence.  
 ACCESSION A2747682  
 VERSION A2747682.1 GI:12531762  
 KEYWORDS GSS.  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 565)  
 AUTHORS Zhao, S., Niernman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,  
 Russell, D., de Jong, P., and Fraser, C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-24  
 JOURNAL Unpublished  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end  
 page: http://www.tigr.org/tbcbac/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 80 row: N column: 12  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1. 565  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-80N12"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pFARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
 library was cloned in the pFARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."

BASE COUNT 217 a 119 c 106 g 123 t  
 ORIGIN

Query Match 0.9%; Score 81; DB 28; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 1.e-17;  
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4593 GCAGTGTGCGACACAGCCCTTAATCCAGCAGCTTTGGAGGCGAGCAGACATTTCTG 4652  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 83 GCAGTGTGCGACACAGCCCTTAATCCAGCAGCTTTGGAGGCGAGCAGACATTTCTG 142  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 4653 AGTTCAAGCGCCAGCTGTGCT 4673  
 ||||||||||||||||||||  
 Db 143 AGTTCAAGCGCCAGCTGTGCT 163  
 ||||||||||||||||||||

RESULT 8  
 BB813805/c 373 bp mRNA linear EST 19-NOV-2001  
 LOCUS BB813805  
 DEFINITION BB813805 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus  
 musculus cDNA clone G730020N16 3', mRNA sequence.  
 ACCESSION BB813805  
 VERSION BB813805.1 GI:16986434  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 373)  
 AUTHORS Akimura,T., Atakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirazane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,K., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)  
 JOURNAL Unpublished  
 COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gscl.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

FEATURES  
 source  
 location/Qualifiers  
 1..373  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="G730020N16"  
 /tissue\_type="lung"  
 /cell\_line="RCB-0558 LLC"  
 /clone\_lib="RIKEN full-length enriched, lung RCB-0558 LLC cDNA"

BASE COUNT  
 ORIGIN  
 99 a 83 c 72 g 119 t

Query Match 0.9%: Score 80; DB 10; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-17;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2202 GAAGTCTGAGTTCATCTAGCAACACATGCTGCTACACCATCATATGAGA 2261  
 |||||||  
 DB 140 GAAGTCTGAGTTCATCTAGCAACACATGCTGCTACACCATCATATGAGA 81  
 |||||||  
 Y 2262 TCTGACACCCCTCTTCTGTG 2281  
 |||||||  
 DB 80 TCTGACACCCCTCTTCTGTG 61  
 |||||||

RESULT 9  
 A2902196 713 bp DNA linear GSS 05-MAR-2001  
 LOCUS RPCI-24-158J11.TJ RPCI-24 Mus musculus genomic clone RPCI-24-158J11  
 DEFINITION , genomic survey sequence.  
 ACCESSION A2902196

VERSION A2902196.1 GI:13221141  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Eukaryota: Euteleia: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
 1 (bases 1 to 713)  
 AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintre,B., Levins,M., Tseaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 JOURNAL Unpublished  
 COMMENT Other GSSs: RPCI-24-158J11.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac-ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html)  
 plate: 158 row: J column: 11  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 location/Qualifiers  
 1..713  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-158J11"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pTRABAC1; Site:1: BamHI; Site:2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT  
 ORIGIN  
 230 a 154 c 165 g 164 t

Query Match 0.9%: Score 79; DB 28; Length 713;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-17;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 4596 GTGGTGACACAGCGCTTAATCCAGCATTGGAGAGCAGACGATTTCTGAGT 4655  
 |||||||  
 DB 396 GTGGTGACACAGCGCTTAATCCAGCATTGGAGAGCAGACGATTTCTGAGT 455  
 |||||||  
 Y 4656 TCAAGGCCAGCTGTGCTA 4674  
 |||||||  
 DB 456 TCAAGGCCAGCTGTGCTA 474  
 |||||||

RESULT 10  
 BB662563 671 bp mRNA linear EST 26-OCT-2001  
 LOCUS BB662563 RIKEN full-length enriched, 15 days embryo head Mus  
 DEFINITION musculus cDNA clone D930008C23 5', mRNA sequence.  
 ACCESSION BB662563  
 VERSION BB662563.1 GI:16496317  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Euteleia: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
 1 (bases 1 to 671)  
 AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

TITLE  
JOURNAL  
COMMENT

Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koye, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arikawa, T., et al. 2001)  
Unpublished

contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.

FEATURES  
source

Location/Qualifiers  
1. 671  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="D930008C23"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="15 days embryo"  
/lab\_host="DH10B"  
/clone\_idb="RIKEN full-length enriched, 15 days embryo head"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGATCTCGAGTAAATTAATTCACCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGATCTCGAGTAAATTAATTCACCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda  
FLC I"

BASE COUNT  
ORIGIN

153 a 147 c 150 g 221 t

Query Match 0.8%; Score 76; DB 10; Length 671;  
Best Local Similarity 100.0%; Pred. No. 5.7e-16;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4638 GCAGGAGATTTCTGAGTTCAGGCCAGCGCTGCTATAGAGTTCAGTCCAGACAGCA 4697  
|||||  
Db 286 GCGAGGAGATTTCTGAGTTCAGGCCAGCGCTGCTATAGAGTTCAGTCCAGACAGCA 227  
|||||  
QY 4698 GGGCTACACAGAGAAA 4713  
|||||  
Db 226 GGGCTACACAGAGAAA 211  
|||||

RESULT 11  
AK086147/c 1083 bp mRNA linear HTC 05-DEC-2002  
LOCUS  
DEFINITION  
Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:ID930008C23 product:unclassified, full insert sequence.  
ACCESSION  
AK086147.1 GI:26103252  
VERSION  
AK086147.1  
KEYWORDS  
HTC; CAP Trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
AUTHORS  
TITLE  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
MEDLINE  
99279253  
PUBMED  
10349636  
REFERENCE  
AUTHORS  
TITLE  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
MEDLINE  
20499374  
PUBMED  
11042159  
REFERENCE  
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL  
MEDLINE  
20530913  
PUBMED  
11076861  
REFERENCE  
AUTHORS

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakai, F., Pesole, G., Quackenbush, J., Schirml, L. M., Stabaldi, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M., F. Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Welt, C., Whitaker, C., Winding, L., Wynshaw-Boris, A., Yoshihide, K., Hasegawa, Y., Kawai, H., Kohlsukl, S.

TITLE and Hayashizaki, Y.  
JOURNAL Functional annotation of a full-length mouse cDNA collection  
MEDLINE Nature 409 (6821), 685-690 (2001)  
PUBMED 21085660  
11217851

REFERENCE  
AUTHORS 5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE 6 (bases 1 to 1083)  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

COMMENT  
TITLE Direct Submission  
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT  
TITLE cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers

FEATURES  
source 1..1083  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:D93008C23"  
/db\_xref="taxon:10090"  
/clone="D93008C23"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="15 days embryo"  
misc\_feature 1..1083  
/note="unclassified"

BASE COUNT 247 a 239 c 241 g 356 t

ORIGIN  
Query Match 0.88; Score 76; DB 11; Length 1083;  
Best Local Similarity 100.0%; Pred. No. 4,9e-16;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4638 GCAGCGAGATTTCGATTCAGGCGCGCTGTATAGAGTTCAGACACCA 4697  
|||||  
DB 286 GCAGCGAGATTTCGATTCAGGCGCGCTGTATAGAGTTCAGACACCA 227  
|||||

QY 4698 GGGCTACACAGAGAAA 4713  
|||||  
DB 226 GGGCTACACAGAGAAA 211  
|||||

RESULT 12  
CD538197/c 438 bp mRNA linear EST 11-JUN-2003  
LOCUS B0201B07-3 NIA Mouse Embryonic Germ Cell cDNA Library (long) Mus

ACCESSION musculus cDNA clone NIA:B0201B07 IMAGE:30103506 3', mRNA sequence.  
CD538197  
VERSION CD538197.1 GI:31585932  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 438)  
AUTHORS Piao, Y., Ko, N.-T., Lim, M.-K. and Ko, M.-S.-H.  
TITLE Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method  
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)  
MEDLINE 21429098  
PUBMED 11544199

COMMENT  
JOURNAL Contact: Dawood B. Dudekula  
MEDLINE Laboratory of Genetics  
PUBMED National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: B0201 Row: B Column: 07  
Seq primer: -21M13 Forward  
High quality sequence stop: 438  
POLYA-Yes.

FEATURES  
source 1..438  
Location/Qualifiers  
1..438  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="nlaEST:B0201B07-3"  
/db\_xref="taxon:10090"  
/clone="NIA:B0201B07 IMAGE:30103506"  
/sex="male"  
/dev\_stage="embryonic day 8"  
/lab\_host="DH10b"  
/clone\_lib="NIA Mouse Embryonic Germ Cell cDNA Library (long)"  
/note="Vector: PCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199]. Total RNAs were obtained from Dr. Mark G. Carter (NIH/NIA-IRP). EG cells were cultured at 37. C, 5% CO2 in DMEM supplemented with 15% ES cell qualified FBS, 0.1mM non-essential amino acids, 2 mM glutamine, penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM beta-mercaptoethanol, and 10000000 units of LIF per liter. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]: 5'-PGACTAGTCTAGATCCGAGCGCGCCCTTTTTTT-3' from 2.5 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker Lf-SalI, purified by phenol/chloroform, and separated from free linkers by Genticon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Genticon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of PCMV-SPORT6 plasmid vector. The DH10b E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 4.0 kb. The library was constructed by Yulan Piao."

BASE COUNT 118 a 88 c 100 g 132 t

ORIGIN  
Query Match 0.88; Score 75; DB 14; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.5e-15;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY	4645	GATTCTTCAGTTCACAGCCACCTGCGTCTATAGAGTGAGTTCACAGGACAGCCAGGGCTAC	4704
DB	79	GATTCTTCAGTTCACAGCCACCTGCGTCTATAGAGTGAGTTCACAGGACAGCCAGGGCTAC	20
QY	4705	ACAGAGAAACCTGCT	4719
DB	19	ACAGAGAAACCTGCT	5
RESULT 13	AZ427733	540 bp DNA linear	GSS 03-OCT-2000
LOCUS	1M0209H24R	Mouse 10kb plasmid	U08C1M library Mus musculus genomic
DEFINITION	clone U08C1M0209H24 R, genomic survey sequence.		
ACCESSION	AZ427733		
VERSION	AZ427733.1	GI:10551746	
KEYWORDS	GSS.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 540)		
AUTHORS	Islam, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamll, C., Duan, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.		
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts		
JOURNAL	unpublished		
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0209 row: H column: 24 Seq primer: CACACAGCAACACGCTATGACC Class: plasmid ends High quality sequence stop: 540.		
FEATURES	Location/Qualifiers		
SOURCE	1..540		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="U08C1M0209H24"		
	/sex="Male"		
	/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"		
	/clone_lib="Mouse 10kb plasmid U08C1M library"		
	/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource ( <a href="http://www.jax.org/resources/documents/dnares/">http://www.jax.org/resources/documents/dnares/</a> ). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1147321141gblAfl29072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."		
BASE COUNT	188 a	113 c	122 g
ORIGIN			116 t 1 others

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:00 ; Search time 525 seconds  
(without alignments)  
7562.364 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995  
Sequence: 1 gatctggtgaccacaagaca.....cttaagctctccaccatg 8995

Scoring table: OLIGO\_NDC  
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	64	0.7	3588	1	US-08-197-792-32
C 2	64	0.7	3588	1	US-08-459-850-32
C 3	64	0.7	3588	1	US-08-459-214-32
C 4	63	0.7	80246	3	US-09-078-294-4
C 5	62	0.7	3381	3	US-09-009-119-1
C 6	62	0.7	3381	4	US-09-371-507-1
C 7	62	0.7	3383	5	PCT-US95-08098-1
C 8	61	0.7	338	4	US-09-495-050A-105
C 9	61	0.7	818	3	US-08-205-697A-4
C 10	61	0.7	818	3	US-08-702-525-4
C 11	61	0.7	818	5	PCT-US95-02576-4
C 12	61	0.7	1570	3	US-08-205-697A-10
C 13	61	0.7	1570	3	US-08-702-525-10
C 14	61	0.7	1570	3	PCT-US95-02576-10
C 15	61	0.7	1606	3	US-08-702-525-64
C 16	61	0.7	1606	4	US-09-645-926A-5
C 17	61	0.7	1816	4	US-08-702-525-1
C 18	61	0.7	1888	3	US-08-205-697A-1
C 19	61	0.7	1888	3	US-08-702-525-1
C 20	61	0.7	1888	5	PCT-US95-02576-1
C 21	61	0.7	2516	3	US-08-205-697A-3
C 22	61	0.7	2516	3	US-08-702-525-3
C 23	61	0.7	2516	5	PCT-US95-02576-3
C 24	61	0.7	2529	3	US-09-051-969A-5
C 25	61	0.7	4157	2	US-08-871-266A-1
C 26	61	0.7	4157	2	US-08-819-458A-1
C 27	61	0.7	4157	2	US-09-018-864A-1

C 28	61	0.7	4157	3	US-08-871-267B-1	Sequence 1, Appl1
C 29	61	0.7	4157	3	US-09-618-419-1	Sequence 1, Appl1
C 30	61	0.7	55298	4	US-09-491-356C-1	Sequence 1, Appl1
C 31	61	0.7	80595	3	US-09-078-294-3	Sequence 3, Appl1
C 32	59	0.7	223	1	US-08-222-177A-14	Sequence 14, Appl1
C 33	59	0.7	1920	4	US-09-534-638-6	Sequence 6, Appl1
C 34	59	0.7	9840	4	US-09-534-638-1	Sequence 1, Appl1
C 35	57	0.6	57	1	US-08-222-177A-128	Sequence 128, Appl1
C 36	57	0.6	113	4	US-09-354-147C-36	Sequence 36, Appl1
C 37	56	0.6	747	4	US-09-220-132-92	Sequence 92, Appl1
C 38	56	0.6	2775	1	US-08-149-096A-1	Sequence 1, Appl1
C 39	56	0.6	63588	4	US-09-873-404-3	Sequence 3, Appl1
C 40	55	0.6	211	4	US-09-354-147C-29	Sequence 29, Appl1
C 41	55	0.6	10409	3	US-08-772-440-33	Sequence 33, Appl1
C 42	55	0.6	176373	3	US-09-128-155-17	Sequence 17, Appl1
C 43	54	0.6	2121	4	US-09-425-488-1	Sequence 1, Appl1
C 44	54	0.6	2235	3	US-09-334-601-3	Sequence 3, Appl1
C 45	54	0.6	2360	3	US-08-916-043-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-197-792-32/c  
; Sequence 32, Application US/08197792  
; Patent No. 5525488  
GENERAL INFORMATION:  
; APPLICANT: Anthony J. Mason  
; APPLICANT: Peter H. Seeburg  
; TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or Beta Chains of Inhibin a  
; NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/197,792  
; FILING DATE: 16-FEB-1994  
CLASSIFICATION:  
; CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/958414  
; FILING DATE: 08-OCT-1992  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/744207  
; FILING DATE: 12-AUG-1991  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/215466  
; FILING DATE: 05-JUL-1988  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 06/906729  
; FILING DATE: 31-DEC-1986  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 06/827710  
; FILING DATE: 07-FEB-1986  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 06/783910  
; FILING DATE: 03-OCT-1985  
AUTORNEY/AGENT INFORMATION:  
; NAME: Hasak, Janet E.  
; REGISTRATION NUMBER: 28, 616  
; REFERENCE/DOCKET NUMBER: 297P2D4  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1896  
; TELEFAX: 415/952-9881



```

: FILING DATE: 03-OCT-1985
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 297P2D6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3588 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-459-214-32

```

[illegible]

```

: RESULT 4
: US-09-078-294-4/c
: Sequence 4, Application US/09078294
: Patent No. 6265211
: GENERAL INFORMATION:
: APPLICANT: Choo, Kong-Hong Andy
: APPLICANT: Du Sart, Desiree
: APPLICANT: Cancilla, Michael R.
: TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
: FILE REFERENCE: Davies Col
: CURRENT APPLICATION NUMBER: US/09/078,294
: CURRENT FILING DATE: 1998-05-13
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 80246
: TYPE: DNA
: ORGANISM: Nucleotide sequence of NC-contlg
: US-09-078-294-4

```

[illegible]

```

RESULT 5
US-09-009-119-1
; Sequence 1, Application US/09009119
; Patent No. 6160206
; GENERAL INFORMATION:
; APPLICANT: SATO, RYO
; APPLICANT: BOYNTON, John E.
; APPLICANT: GILHAM, Nicholas W.
; APPLICANT: HARRIS, Elizabeth H.
; TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene

```

```

: FILE REFERENCE: substitute sequence listing
: CURRENT APPLICATION NUMBER: US/09/009,119
: CURRENT FILING DATE: 1996-01-20
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3381
: TYPE: DNA
: ORGANISM: Chlamydomonas reinhardtii
: US-09-009-119-1

```

[illegible]

```

RESULT 6
US-09-371-507-1
: Sequence 1, Application US/09371507
: Patent No. 6346656
: GENERAL INFORMATION:
: APPLICANT: SATO, RYO
: APPLICANT: BOYNTON, John E.
: APPLICANT: GILLHAM, Nicholas W.
: APPLICANT: HARRIS, Elizabeth H.
: TITLE OF INVENTION: Porphyryn Accumulating-Type Herbicide Resistance Gene
: FILE REFERENCE: substitute sequence listing
: CURRENT APPLICATION NUMBER: US/09/371,507
: PRIOR FILING DATE: 1999-08-18
: PRIOR APPLICATION NUMBER: 09/009,119
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3381
: TYPE: DNA
: ORGANISM: Chlamydomonas reinhardtii
US-09-371-507-1

```

[illegible]

```

RESULT 7
PCT-US95-09098-1
: Sequence 1, Application PC/TUS9509098
: GENERAL INFORMATION:
: APPLICANT: Sato, Ryo
: APPLICANT: Boynton, John E.
: APPLICANT: Gillham, Nicholas W.
: APPLICANT: Harris, Elizabeth H.
: TITLE OF INVENTION: Porphyrin-Accumulating Type Herbicide
: TITLE OF INVENTION: Resistance Gene
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

```

[illegible][illegible]











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:00 ; Search time 2039 Seconds  
(without alignments)  
11908.507 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995  
Sequence: 1 gatctgtgtcaccacagagaca.....ctctagctctccaccatg 8995

Scoring table: OLIGO\_NDC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8995	100.0	8995	21	AAA48515 Murine villin gene
2	8631	96.2	8993	21	AAA49605 Mouse villin gene
3	66	0.7	2628	24	AB519044 Mouse WU-1 haemato
4	65	0.7	1623	24	AB565621 Mouse fabe gene (E
5	64	0.7	574	21	ABN81091 Shrimp polynucleot
6	64	0.7	2447	22	AA161087 Human polynucleoti
7	63	0.7	80	23	AA513715 Simple sequence re
8	63	0.7	279	21	AAA74225 Lobliolly pine SSR

9	63	0.7	382	15	AA055246 Grapevine ribosoma
10	63	0.7	382	21	AAA74316 Lobliolly pine SSR
11	63	0.7	414	22	AA193748 Human polynucleoti
12	63	0.7	4580	17	AA132034 Proliferation-inhi
13	63	0.7	80240	20	AA183940 NC-ctrlg derived
14	63	0.7	219	13	AA027094 Fragile X diagnost
15	62	0.7	381	24	ABX04909 Conus sp conotoxin
16	62	0.7	463	21	ABN81124 Shrimp polynucleot
17	62	0.7	2573	19	AA162927 C. reinhardtii pro
18	62	0.7	3316	23	AB110742 Drosophila melanog
19	62	0.7	3381	18	AA178597 Porphyritin accumula
20	62	0.7	3381	19	AA162928 C. reinhardtii pro
21	62	0.7	3383	18	AA160965 Porphyritin-accumula
22	62	0.7	14704	13	AA020685 PKS 741 insert con
23	62	0.7	14704	13	AA020685 PKS 741 insert con
24	62	0.7	42999	24	AB565032 Invertebrate forag
25	62	0.7	173810	24	ABN85752 Mouse chromosome 1
26	61	0.7	66	23	AA513714 Simple sequence re
27	61	0.7	68	21	AA298506 H. discus derived
28	61	0.7	76	23	AA513704 Simple sequence re
29	61	0.7	82	21	AA298483 H. discus derived
30	61	0.7	90	23	AA513709 Simple sequence re
31	61	0.7	127	25	AB219796 Group III CDNA can
32	61	0.7	140	22	AAK90823 Human digestive sy
33	61	0.7	219	21	ABN81161 Shrimp polynucleot
34	61	0.7	266	13	AA027097 Fragile X diagnost
35	61	0.7	282	13	AA027100 Fragile X diagnost
36	61	0.7	316	22	AAH29391 Drosophila melanog
37	61	0.7	362	21	AAA74223 Lobliolly pine SSR
38	61	0.7	407	22	AA166146 Novel human polynu
39	61	0.7	448	18	AA168682 Penaeus vannamei g
40	61	0.7	475	22	AB443925 Human breast cell
41	61	0.7	475	22	AB443925 Human foetal liver
42	61	0.7	475	22	AB443925 Human foetal liver
43	61	0.7	475	22	AAK02673 Probe #2623 for ge
44	61	0.7	475	22	AAK28111 Human brain expres
45	61	0.7	475	22	AA112688 Human bone marrow
					Probe #2621 for ge

ALIGNMENTS

RESULT 1	
AAA48515	AAA48515 standard; DNA; 8995 BP.
XX	XX
AC	AAA48515;
XX	XX
DT	19-DEC-2000 (first entry)
XX	XX
DE	Murine villin gene regulatory region.
XX	XX
KW	Mouse; villin; regulatory region; digestive tract;
KW	colorectal cancer mouse model; ds.
XX	XX
OS	Mus sp.
XX	XX
FH	FH
FT	Key
FT	location/Qualifiers
FT	3442
FT	/*tag= a
FT	/note= "transcriptional start site"
FT	3442..3487
FT	/*tag= b
FT	/number= 1
FT	3488..8981
FT	/*tag= c
FT	/number= 1
FT	8982..8995
FT	/*tag= d
FT	/number= 2
FT	8993
FT	/*tag= e
FT	/note= "translational start site"

xx	WO200034492-A1.
PN	
xx	
PD	15-JUN-2000.
xx	
xx	
PF	09-DEC-1998; 98WO-EP08009.
xx	
PF	09-DEC-1998; 98WO-EP08009.
xx	
PA	(CNRS ) CENT NAT RECH SCI.
xx	
PA	(CURIE-) INST CURIE.
xx	
PI	Pinto D, Robine S, Jaisser F, Louvard D;
xx	
DR	WPI; 2000-42343/36.
xx	
PT	Novel nucleotide sequence derived from mouse villin gene for targeted
xx	
PT	expression of transgenes in immature and differentiated epithelial
xx	
PT	cells of intestine or urogenital tracts -
xx	
PS	Claim 3; Fig 6; 54pp: English.
xx	
CC	The present sequence comprises the murine villin gene regulatory
xx	
CC	region, the first intron and part of the first two exons. It has been
xx	
CC	shown that this region directs the expression of the villin gene in the
xx	
CC	intestine and urogenital tracts, and thus could be used in a fusion
xx	
CC	gene to direct expression of exogenous genes in these areas. This could
xx	
CC	be used, for example, to create a mouse model for colorectal cancer.
xx	
SO	Sequence 8995 BP; 2275 A; 2105 C; 2258 G; 2357 T; 0 other;

Oy	1661	TTAGCATGCTGCTGAGAAGATAGAAAAATGAAGACTTGAAGTCTAGTCTGGAACCCACA	1740
Db	1661	TTAGCATGCTGCTGAGAGAAGATAGAAAAATGAACCTTAGACTCTAGTCTGGAACCCACA	1740
Oy	1741	GAGGAGGCGGAAGAACCCCACTCCTGAAGTGTCTCTGAGAGCTTCACATACAACTTCACAT	1800
Db	1741	GAGGAGGCGGAAGAACCCCACTCCTGAAGTGTCTCTGAGAGCTTCACATACAACTTCACAT	1800
Oy	1801	AATAGTTCATATGATTAATTAATTAATTAATTCCTTTTAAAGATATATGTTGGAGAGGA	1860
Db	1801	AATAGTTCATATGATTAATTAATTAATTAATTCCTTTTAAAGATATATGTTGGAGAGGA	1860
Oy	1861	GAGATGCTCAGCTTCACGAGACACTGCTGCTCTTGAGAGAGACTAGATTCAGTTCCTC	1920
Db	1861	GAGATGCTCAGCTTCACGAGAGACTGCTGCTCTTGAGAGAGACTAGATTCAGTTCCTC	1920
Oy	1921	AGGACTCATATGAGTGGCTCACAGGCACATCTGTAAATCCAGTTCGAGAGGCTTCCACACCT	1980
Db	1921	AGGACTCATATGAGTGGCTCACAGGCACATCTGTAAATCCAGTTCGAGAGGCTTCCACACCT	1980
Oy	1981	CTTCTGGCCTCCACAGGCACACATACATACAGACACATACATGAGAGCAAAACAC	2040
Db	1981	CTTCTGGCCTCCACAGGCACACATACATACAGACACATACATGAGAGCAAAACAC	2040
Oy	2041	CATACACATTAATTAATTAATTAAGAACTTAAAGGTGCATGTGTTGTTAAACATTGCT	2100
Db	2041	CATACACATTAATTAATTAATTAAGAACTTAAAGGTGCATGTGTTGTTAAACATTGCT	2100
Oy	2101	TACACATCTCATTTGAAGACATGTATCAACGACACACTGAAGAGGATCTGGGCTGGAG	2160
Db	2101	TACACATCTCATTTGAAGACATGTATCAACGACACACTGAAGAGGATCTGGGCTGGAG	2160
Oy	2161	AGATGGCTCAGCGGTTTAAAGACACTGATGCTGCTCCGAAAGAAAGTCTGAGTTCAAT	2220
Db	2161	AGATGGCTCAGCGGTTTAAAGACACTGATGCTGCTCCGAAAGAAAGTCTGAGTTCAAT	2220
Oy	2221	CCTAGCAACACATGCTGGCTCACACCATCCATCATATAGATCTGACACCTCTTCTGCT	2280
Db	2221	CCTAGCAACACATGCTGGCTCACACCATCCATCATATAGATCTGACACCTCTTCTGCT	2280
Oy	2281	GCATCTGAAGACAGCTGCAGAGCTACATGCTACTAGTATATACATTAATTAATTCCTTT	2340
Db	2281	GCATCTGAAGACAGCTGCAGAGCTACATGCTACTAGTATATACATTAATTAATTCCTTT	2340
Oy	2341	TTTAAAAAAATGAAGAGGATCTGAGACCTCAAAAGAGATTTATGACAGTGACTCAG	2400
Db	2341	TTTAAAAAAATGAAGAGGATCTGAGACCTCAAAAGAGATTTATGACAGTGACTCAG	2400
Oy	2401	GGTGATTAATCTATCTCTGGAGTTTTTCCCTTCCGCTTGCCACTGGAAGTGGAGACAG	2460
Db	2401	GGTGATTAATCTATCTCTGGAGTTTTTCCGCTTGCCACTGGAAGTGGAGACAG	2460
Oy	2461	CCCCCTTTCATTCACAAAGCGGGGCTACATTTATTTTGAAACAAACAGCACTGCACT	2520
Db	2461	CCCCCTTTCATTCACAAAGCGGGGCTACATTTATTTTGAAACAAACAGCACTGCACT	2520
Oy	2521	ATGTTTACTGTCCTTGCTAGCTATGAGCAGCGCAGCGCGCGGCACACACACACAC	2580
Db	2521	ATGTTTACTGTCCTTGCTAGCTATGAGCAGCGCAGCGCGCGGCACACACACACAC	2580
Oy	2581	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGGAAGTCA	2640
Db	2581	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGGAAGTCA	2640
Oy	2641	AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTTAAAGAGACACAGATTCC	2700
Db	2641	AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTTAAAGAGACACAGATTCC	2700
Oy	2701	AAGGTGCGAGAAGATCTACAGGGGCGAGGCGAGGAGGGGGAAGCGCCATGGTTTTCC	2760
Db	2701	AAGGTGCGAGAAGATCTACAGGGGCGAGGCGAGGAGGGGGAAGCGCCATGGTTTTCC	2760

QY	2761	AGAGACCTACAGCAGAGGGCAGCAAGGCAGATCCCCAGGTCCAGGCGCAGGGAGGTGGAGG	2820
Db	2761	AGAGACCTACAGCAGAGGGCAGCAAGGCAGATCCCCAGGTCCAGGCGCAGGGAGGTGGAGG	2820
QY	2821	CCCTGTGTTCCAGAGGAAGGCGAGGCGGAGAACAGGGCTCCAAAGGCACAGGTTATGGCA	2880
Db	2821	CCCTGTGTTCCAGAGGAAGGCGAGGCGGAGAACAGGGTTCAAAGGCACAGGTTATGGCA	2880
QY	2881	GCTCATATAAAGTGGAGGCGTGGCTCACTCAGAAAGGAGGAAGAAAGGGAAAGGCCCTTGT	2940
Db	2881	GCTCATATAAAGTGGAGGCGTGGCTCACTCAGAAAGGAGGAAGAAAGGGAAAGGCCCTTGT	2940
QY	2941	GCCCACTGAGCGAGGGGTCTATGTAAGAGAGATCTGCGAGGGGTGCCAGAGACCCAC	3000
Db	2941	GCCCACTGAGCGAGGGGTCTATGTAAGAGAGATCTGCGAGGGGTGCCAGAGACCCAC	3000
QY	3001	CTGTCTGTGCCAAGGGAACCCCAAGTGTGAAGCTGTGGGCTTGGGTGCTGAATTCACACTA	3060
Db	3001	CTGTCTGTGCCAAGGGAACCCCAAGTGTGAAGCTGTGGGCTTGGGTGCTGAATTCACACTA	3060
QY	3061	CAAGACCCCAAGAGGCTCTACTCCATCCCAATCCACAGTGGCCCCCTGGCCCCGACACCCCA	3120
Db	3061	CAAGACCCCAAGAGGCTCTACTCCATCCCAATCCACAGTGGCCCCCTGGCCCCGACACCCCA	3120
QY	3121	CCCCCGACTCCCGGTCCACTCTCTAGGGCTGAGAGGGTGGCCAGCCCTGGTGGGGTTCG	3180
Db	3121	CCCCCGACTCCCGGTCCACTCTCTAGGGCTGAGAGGGTGGCCAGCCCTGGTGGGGTTCG	3180
QY	3181	CTACCTGAGGTTAAGCCCAAGTCTCTAGCGGGAAGTGCACCCCAATCCCTGAAGTGGAGA	3240
Db	3181	CTACCTGAGGTTAAGCCCAAGTCTCTAGCGGGAAGTGCACCCCAATCCCTGAAGTGGAGA	3240
QY	3241	GCCAAAGGCGGGGGACACAGGCGAGCTGTAGAGCTGTGGCTGGGGCTCTAGGTTCC	3300
Db	3241	GCCAAAGGCGGGGGACACAGGCGAGCTGTAGAGCTGTGGCTGGGGCTCTAGGTTCC	3300
QY	3301	AGGACCTGGGCACCTACTTCCCAACCCCCCATCATCTCTCTGGGGCCCTATCTTCC	3360
Db	3301	AGGACCTGGGCACCTACTTCCCAACCCCCCATCATCTCTCTGGGGCCCTATCTTCC	3360
QY	3361	CTTATATGTAAGGAAGTCTCTGGGGGGGGGGGGTGGTGGAGCAAAAGATCTGTCG	3420
Db	3361	CTTATATGTAAGGAAGTCTCTGGGGGGGGGGGGTGGTGGAGCAAAAGATCTGTCG	3420
QY	3421	GTCCTCTGACACCAGCTTGGCCCAACTTCTCTAAGTCTTCCAGAGTGTGGCTGCCCTTTC	3480
Db	3421	GTCCTCTGACACCAGCTTGGCCCAACTTCTCTAAGTCTTCCAGAGTGTGGCTGCCCTTTC	3480
QY	3481	CAGACAGGTAAGGCATTTGGTGGGGACACATGTTGACACACAGGTGTTGGAGGGACAG	3540
Db	3481	CAGACAGGTAAGGCATTTGGTGGGGACACATGTTGACACACAGGTGTTGGAGGGACAG	3540
QY	3541	GGTCCCTGCTCTCTCTGGGACCCCTGTGCTTCTGTAGACACTGTGGATTAAGTTTGGGGG	3600
Db	3541	GGTCCCTGCTCTCTCTGGGACCCCTGTGCTTCTGTAGACACTGTGGATTAAGTTTGGGGG	3600
QY	3601	TGAGGTAGAGGCTCTGAAACTCTGAAAGAAAGCAAGACAGCAGGCTGTCTTGGGCT	3660
Db	3601	TGAGGTAGAGGCTCTGAAACTCTGAAAGAAAGCAAGACAGCAGGCTGTCTTGGGCT	3660
QY	3661	TCATATGAAGAATTCACAGACCCCTTCTCTGTAGTCACTTCCTGCTCATCTGTGTAG	3720
Db	3661	TCATATGAAGAATTCACAGACCCCTTCTCTGTAGTCACTTCCTGCTCATCTGTGTAG	3720
QY	3721	ATTCCTGAGGACCAAGGTGGCTCTCTGGGACCTCAATTTCTAACAATTAACACAGACAGT	3780
Db	3721	ATTCCTGAGGACCAAGGTGGCTCTCTGGGACCTCAATTTCTAACAATTAACACAGACAGT	3780
QY	3781	CCTGAGACTTGGACCTCGCTGCTATTTACTACTTCTCTGTGGCTGCTCATTTCTGTGT	3840
Db	3781	CCTGAGACTTGGACCTCGCTGCTATTTACTACTTCTCTGTGGCTGCTCATTTCTGTGT	3840
QY	3841	TCATGTCTTACACATCTGAAATGGTCTTCTTGTGTCAACATTTCCCTGACACTCTCTGGGA	3900

```
Db 3841 TCACTGCTTACACATCTGAAATGGTTCTTGTGTGCACATTCCTCCCTACACATCTCCGTGGA 3900
Qy 3901 GGTCTATCTCTGGACATGTATCTCTGGATGTAACTGACAGCCACGAGAGAGGGG 3960
Db 3901 GGTCTATCTCTGGACATGTATCTCTGGATGTAACTGACAGCCACGAGAGAGGGG 3960
Qy 3961 AGAGTCAGAGAGCTGTCTTAGGCGCTTATAGGCGTGCACATCACCCCTTCTCTGAAT 4020
Db 3961 AGAGTCAGAGAGCTGTCTTAGGCGCTTATAGGCGTGCACATCACCCCTTCTCTGAAT 4020
Qy 4021 GGGCCCTCATTTTTCGGTTACCATGATCTATTTATATCAGAGTGGGACGTGAAGCA 4080
Db 4021 GGGCCCTCATTTTTCGGTTACCATGATCTATTTATATCAGAGTGGGACGTGAAGCA 4080
Qy 4081 AACCTGCCAGAAAGTTGGAGACTCACTCAGACCAAGTTATCTGCTCAGAAATCCCGTG 4140
Db 4081 AACCTGCCAGAAAGTTGGAGACTCACTCAGACCAAGTTATCTGCTCAGAAATCCCGTG 4140
Qy 4141 TCACCTTGAAGTGGAGAAATCTGCTTGGGGGCTTCAGAGTCTTGTTAGCAGAGAGGT 4200
Db 4141 TCACCTTGAAGTGGAGAAATCTGCTTGGGGGCTTCAGAGTCTTGTTAGCAGAGAGGT 4200
Qy 4201 ATCCTTGTATAGGGCAGTACCTAGTCTATGTTTACTACTTCCCTGTCAGTTAAAG 4260
Db 4201 ATCCTTGTATAGGGCAGTACCTAGTCTATGTTTACTACTTCCCTGTCAGTTAAAG 4260
Qy 4261 CTGGAACATAAAACCCAGCGAGCGCCAGAGATTCTCTACAGTTGTACCCCAAGAACAA 4320
Db 4261 CTGGAACATAAAACCCAGCGAGCGCCAGAGATTCTCTACAGTTGTACCCCAAGAACAA 4320
Qy 4321 AGACAGTATGATATGCAAGGATAGTACTGGGAGAGAGAACTTAAACCCCAAG 4380
Db 4321 AGACAGTATGATATGCAAGGATAGTACTGGGAGAGAGAACTTAAACCCCAAG 4380
Qy 4381 GCGCCAGAGTCCGTTCCCTAGTTGACAATGCCATATGAGTGTAGTACTAGTGGGCTG 4440
Db 4381 GCGCCAGAGTCCGTTCCCTAGTTGACAATGCCATATGAGTGTAGTACTAGTGGGCTG 4440
Qy 4441 TGAGTTGGTAGCTACAGCATGATGATGTTTATGTTGAGTGTATGTTATATCTAGAC 4500
Db 4441 TGAGTTGGTAGCTACAGCATGATGATGTTTATGTTGAGTGTATGTTATATCTAGAC 4500
Qy 4501 TTGGAGAGCTGAAGCAGAGAGATTGCTATATGTTTGAAGCCAGCTGAGCTATAGAGCA 4560
Db 4501 TTGGAGAGCTGAAGCAGAGAGATTGCTATATGTTTGAAGCCAGCTGAGCTATAGAGCA 4560
Qy 4561 GACTTTGTCTTAAAGAAAAATGAAGCCAGCGAGTGTGSCAACCCCTTATGCCA 4620
Db 4561 GACTTTGTCTTAAAGAAAAATGAAGCCAGCGAGTGTGSCAACCCCTTATGCCA 4620
Qy 4621 GCACCTTGGAGGACAGAGCAGAGATTGCTGATTTCAAGGCGACCTGCTATAGAGT 4680
Db 4621 GCACCTTGGAGGACAGAGCAGAGATTGCTGATTTCAAGGCGACCTGCTATAGAGT 4680
Qy 4681 GAGTTCCAGGACAGCGAGGCTTACACAGAGAAACCCCTGTTTGAAGAAACAGAAAAACA 4740
Db 4681 GAGTTCCAGGACAGCGAGGCTTACACAGAGAAACCCCTGTTTGAAGAAACAGAAAAACA 4740
Qy 4741 AACAAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 4800
Db 4741 AACAAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 4800
Qy 4801 TAGGCTGTGTCTGTCTAGTGGTAGAGTTTGGGAGCTTCAGACTTATATTAATAGGCC 4860
Db 4801 TAGGCTGTGTCTGTCTAGTGGTAGAGTTTGGGAGCTTCAGACTTATATTAATAGGCC 4860
Qy 4861 TTTTATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Db 4861 TTTTATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy 4921 AAGTACTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980
Db 4921 AAGTACTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980
```

QY	6061	1GACCTTCCACTTATTAGTACTACATAGGCCAAGGACAGTCCACAGAGGGGTGACTGGGTCTTA	6120
Db	6061	TGACCTTCCACTTATTAGTACTACATAGGCCAAGGACAGTCCACAGAGGGGTGACTGGGTCTTA	6120
QY	6121	CTGAGCTTGGAGGACGGCAGTGGGAGAAATGGGTGACCTCCATTCGATGAGACAGGGCTGAG	6180
Db	6121	CTGAGCTTGGAGGACGGCAGTGGGAGAAATGGGTGACCTCCATTCGATGAGACAGGGCTGAG	6180
QY	6181	CACCAACGATACAAAGTGTTCCTCTGTCTCATGCGACAGGATTCCTGGCCACTTTTCAAG	6240
Db	6181	CACCAACGATACAAAGTGTTCCTCTGTCTCATGCGACAGGATTCCTGGCCACTTTTCAAG	6240
QY	6241	GACTTAGGACTCATCTCTGTGTGGAAACAATATCCAAAGCCCTAAGCCCATTTGGTCT	6300
Db	6241	GACTTAGGACTCATCTCTGTGTGGAAACAATATCCAAAGCCCTAAGCCCATTTGGTCT	6300
QY	6301	AATTAAATCAACAACCCCTGGGGATGACAGGCTCTGAGACGACGAGACCTTTTAAAGATC	6360
Db	6301	AATTAAATCAACAACCCCTGGGGATGACAGGCTCTGAGACGACGAGACCTTTTAAAGATC	6360
QY	6361	CCAGGTGATTTGTATCAGCAGCTGGAAACAACACAGCTACAGGTTCAACAGAAAGAGGC	6420
Db	6361	CCAGGTGATTTGTATCAGCAGCTGGAAACAACACAGCTACAGGTTCAACAGAAAGAGGC	6420
QY	6421	AAAGCTAGGAAAGCTTGGGATGGGGAGCCCTTCCAGGCAAGCACTAAGATGAGAGGCTGTT	6480
Db	6421	AAAGCTAGGAAAGCTTGGGATGGGGAGCCCTTCCAGGCAAGCACTAAGATGAGAGGCTGTT	6480
QY	6481	AGCAGTGTGGCAGCTTCTCTCTGCTGTCATATAGTATCCATCCATCCATCTATCCATCAT	6540
Db	6481	AGCAGTGTGGCAGCTTCTCTCTGCTGTCATATAGTATCCATCCATCCATCCATCTATCCATCAT	6540
QY	6541	ACACCCACCCATTCATTATGACACCCATCCCTTCATCCATCCATCTATCCAGCTATCCACCAC	6600
Db	6541	ACACCCACCCATTCATTATGACACCCATCCCTTCATCCATCCATCTATCCAGCTATCCACCAC	6600
QY	6601	CCAGGCATCCATCCAAACCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT	6660
Db	6601	CCAGGCATCCATCCAAACCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT	6660
QY	6661	CATTATTCACACAGAAAGCTGTAATTGTACTAAATGTGGAGATTTAATTATTTTAGA	6720
Db	6661	CATTATTCACACAGAAAGCTGTAATTGTACTAAATGTGGAGATTTAATTATTTTAGA	6720
QY	6721	AGCTCTGTGATGACTGATGTGCATGATGTGGAGAGATATACCACAGCACAGCAGCTG	6780
Db	6721	AGCTCTGTGATGACTGATGTGCATGATGTGGAGAGATATACCACAGCACAGCAGCTG	6780
QY	6781	TGGCAATCGAAGAAAGTTTGGGTGTTTGTCTCTCCACCGTGTGGGTCTGGGGA	6840
Db	6781	TGGCAATCGAAGAAAGTTTGGGTGTTTGTCTCTCCACCGTGTGGGTCTGGGGA	6840
QY	6841	TTGAACCTCAATTTATNCGGGCTGGTGGGGAAGTGCTTTAACCCGACGATTTTCTGACA	6900
Db	6841	TTGAACCTCAATTTATNCGGGCTGGTGGGGAAGTGCTTTAACCCGACGATTTTCTGACA	6900
QY	6901	CATCATTTATTTATGAAAGACATCTTATGTAGTCCAGGCTGGCCGCAAGCTGTATGTGCG	6960
Db	6901	CATCATTTATTTATGAAAGACATCTTATGTAGTCCAGGCTGGCCGCAAGCTGTATGTGCG	6960
QY	6961	CCAGGAGTACCTTTAACTCCTGCTCTTCCAGGCTCCACCCGAGTGTAGTTTACAGGT	7020
Db	6961	CCAGGAGTACCTTTAACTCCTGCTCTTCCAGGCTCCACCCGAGTGTAGTTTACAGGT	7020
QY	7021	GTTCAACTGTGGAATGCTTTAATCCAGCACTGTGTGGGGGGGGGGGAGGCGGATC	7080
Db	7021	GTTCAACTGTGGAATGCTTTAATCCAGCACTGTGTGGGGGGGGGGGAGGCGGATC	7080
QY	7081	CCTGAGTTGGAGGCCAGTTTGGTCTACAGAGTTTCAGAGATCCTGGGGCTATACAGGAA	7140
Db	7081	CCTGAGTTGGAGGCCAGTTTGGTCTACAGAGTTTCAGAGATCCTGGGGCTATACAGGAA	7140

QY	7141	ACCCATATCCAAACAACAAACAACAAACAAAATAATTTCTGTGCATATATACACAGA	7200
Db	7141	AACCATATCCAAACAACAAACAACAAACAAAATAATTTCTGTGCATATATACACAGA	7200
QY	7201	TTAGAGATATTAGAGGATAGTAGGGCTGGTGGAGGAGATCAATGCTTCTTTGTAAT	7260
Db	7201	TTAGAGATATTAGAGGATAGTAGGGCTGGTGGAGGAGATCAATGCTTCTTTGTAAT	7260
QY	7261	ATAATAGTAAGTACTCTCAGAGATCATATCTATCTATCTATCTATCTATCTATCT	7320
Db	7261	ATAATAGTAAGTACTCTCAGAGATCATATCTATCTATCTATCTATCTATCTATCT	7320
QY	7321	TATCTATCTATCTACCTACCTACCTACCTATCCATCCATCCATCTATCTATAGCCAGG	7380
Db	7321	TATCTATCTATCTACCTACCTACCTACCTATCCATCCATCCATCTATCTATAGCCAGG	7380
QY	7381	CTGCTTTGACCTGAATGCTCCCTATTTCTGGGGTCAACTCTCACCCTAGTGGGTTT	7440
Db	7381	CTGCTTTGACCTGAATGCTCCCTATTTCTGGGGTCAACTCTCACCCTAGTGGGTTT	7440
QY	7441	ACCAACACCCAGACATTTATTTTATTTTATTTTATTTATATCTAGAGACTCAGGGT	7500
Db	7441	ACCAACACCCAGACATTTATTTTATTTTATTTTATTTATTTATATCTAGAGACTCAGGGT	7500
QY	7501	GGACTCAGGGCTTTGTGCATGCTTAAGCAAGCTCTGCGACAGACTGACAGCTCCAGTCC	7560
Db	7501	GGACTCAGGGCTTTGTGCATGCTTAAGCAAGCTCTGCGACAGACTGACAGCTCCAGTCC	7560
QY	7561	CCATTTTGTAGGTGACTCTGTGACAGTTGTCAATTTGCGAGGGCTATGTAGTCTCTC	7620
Db	7561	CCATTTTGTAGGTGACTCTGTGTACAGATTGTCAATTTGCGAGGGCTATGTAGTCTCTC	7620
QY	7621	CACCTCCAGTTCAGACACTTTCTGGTCATCCAGTGGGGGGGCACTGCTGCTACCA	7680
Db	7621	CACCTCCAGTTCAGACACTTTCTGGTCATCCAGTGGGGGGGCACTGCTGCTACCA	7680
QY	7681	GTGCCCTGTCCCTGTCTTGAGACCTACATATTTGCTGTCTGTAAGAGTTCATGTAATG	7740
Db	7681	GTGCCCTGTCCCTGTCTTGAGACCTACATATTTGCGTGTCTGTAAGAGTTCATGTAATG	7740
QY	7741	GGATGCGTCTGTGTATTTCTTTATATGGCTGGCCCCCTTATCTTAGACAGATTGTGTG	7800
Db	7741	GGATGCGTCTGTGTATTTCTTTATATGGCTGGCCCCCTTATCTTAGACAGATTGTGTG	7800
QY	7801	GGCCATGTGTACTGCTATACTCTATCTATCTATCATCATCTATAGGCTTAATAGTGTCCT	7860
Db	7801	GGCCATGTGTACTGCTATACTCTATCTATCTATCATCATCTATAGGCTTAATAGTGTCCT	7860
QY	7861	TGTGTGATAAACCACTTCTGTTCATATTTACTGATGGAAATTTTGCGCCACACCCAC	7920
Db	7861	TGTGTGATAAACCACTTCTGTTCATATTTACTGATGGAAATTTTGCGCCACACCCAC	7920
QY	7921	CCTTTTTTTTTTTATTGAGCAAGGCTTCTGTGTAATCTTGCAATCTTGGCTGTCT	7980
Db	7921	CCTTTTTTTTTTTTTTGAGCAAGGCTTCTGTGTAATCTTGCAATCTTGGCTGTCT	7980
QY	7981	GAGCTCACTGTGTAGACAGGCTGTGAGGGCTGCTTCACCTTTGACACGCTGTAAC	8040
Db	7981	GAGCTCACTGTGTAGACAGGCTGTGAGGGCTGCTTCACCTTTGACACCTGTAAC	8040
QY	8041	AGAGTAGCCATGAACCTTCAAGACAAATTTTCTGTTTGGTTTGTTTTTTACATTTGTGT	8100
Db	8041	AGAGTAGCCATGAACCTTCAAGACAAATTTTCTGTTTGGTTTGTTTTTTACATTTGTGT	8100
QY	8101	TGTATGCGTGTATATGTGCATGTTTGTCTTCACGGTGTACATGTGTACTCTGTGT	8160
Db	8101	TGTATGCGTGTATATGTGCATGTTTGTCTTCACGGTGTCTACATGTGTACTCTGTGT	8160
QY	8161	TGGAGAGAGAAACAACGATGTGCCATCTCTCGATCTACAGCATCTTGTATTAATGTGA	8220
Db	8161	TGGAGAGAGAAACAACGATGTGCCATCTCTCGATCTACAGCATCTTGTATTAATGTGA	8220
QY	8221	TGTATTTATGTATGTTTATTTAGTGTGCCCAAGTATGCAAGTATTTTGTGGAGTTTTCAC	8280

D	b		8221	TGATTATTAAGTATGGTTATTAATTAAGTCGTGCCAAGATATGCAGGATATTTTGTTGAAGTTTTAC	8280
O	y		8281	CTTCCCTTTGTGGGCTCTCCGCAATTAAACTCAGCTCTCGGGCTAGTAGCAATGCCCTTCA	8340
D	b		8281	CTTCCCTTTGTGGGCTCTCCGCAATTAAACTCAGCTCTCGGGCTAGTAGCAATGCCCTTCA	8340
O	y		8341	CTGCATGAGCACATCTCCGCTGCCCTTCCGCTACCCTCTCTTATTTTCCCAGATGGGACTA	8400
D	b		8341	CTGCATGAGCACATCTCCGCTGCCCTTCCGCTACCCTCTCTTATTTTCCCAGATGGGACTA	8400
O	y		8401	CGCACTGCACATGGGCTTAAGCTCACCAAGTCATCATCAGAAGTGGTACGACAGGAGATCTCAG	8460
D	b		8401	CGCACTGCACATGGGCTTAAGCTCACCAAGTCATCATCAGAAGTGGTACGACAGGAGATCTCAG	8460
O	y		8461	GGATATGCTGGCCCTCTGCTCCACAGCTGCTAAGAAATTACAGGCATACATCACTGCTGAAG	8520
D	b		8461	GGATATGCTGGCCCTCTGCTCCACAGCTGCTAAGAAATTACAGGCATACATCACTGCTGAAG	8520
O	y		8521	ATTTTAAACCTGGAATCTCTGAGGATAGAGACGACCTCTACAATGAGAGGCTTTCTTTTGT	8580
D	b		8521	ATTTTAAACCTGGAATCTCTGAGGATAGAGACGACCTCTACAATGAGAGGCTTTCTTTTGT	8580
O	y		8581	GTTTGATTTGGTTTCCCTCTGCATTAAGATACAGGACGTGGAATATAGTAGCTGAGGCTTAC	8640
D	b		8581	GTTTGATTTGGTTTCCCTCTGCATTAAGATACAGGACGTGGAATATAGTAGCTGAGGCTTAC	8640
O	y		8641	ATAACATCTTGTCTCAAAAAAGCCTATAGAGGTAGGAGGTGAGGCTTAAGAAAGGCTT	8700
D	b		8641	ATAACATCTTGTCTCAAAAAAGCCTATAGAGGTAGGAGGTGAGGCTTAAGAAAGGCTT	8700
O	y		8701	AAGCCGGCTGTGATAGCACACAGATAGCCTGCACTATATAGCAAGACCTTTGTTCAAA	8760
D	b		8701	AAGCCGGCTGTGATAGCACACAGATAGCCTGCACTATATAGCAAGACCTTTGTTCAAA	8760
O	y		8761	ACATGAGAGGGGGGTATGTTTAAAGTCTGGGCTGTTATACAGGACCTAAGGAGGCCAA	8820
D	b		8761	ACATGAGAGGGGGGTATGTTTAAAGTCTGGGCTGTTATACAGGACCTAAGGAGGCCAA	8820
O	y		8821	TGTAGACATTTTGACTAAGAAAGATATCATCAAAAGCCGGGTGGGAGGTAAGGTTGG	8880
D	b		8821	TGTAGACATTTTGACTAAGAAAGATATCATCAAAAGCCGGGTGGGAGGTAAGGTTGG	8880
O	y		8881	ACTACAGTGTCAAGAACCCCCCATATGGAAGCCAGTTTCCCTTCTCTGGGCTCAAGC	8940
D	b		8881	ACTACAGTGTCAAGAACCCCCCATATGGAAGCCAGTTTCCCTTCTCTGGGCTCAAGC	8940
O	y		8941	CTGGCTGAGGCGCACGCTCATCAGTCCCTTCCCTCTAGGCGTGCACACATG	8995
D	b		8941	CTGGCTGAGGCGCACGCTCATCAGTCCCTTCCCTCTAGGCGTGCACACATG	8995
<b>RESULT 2</b>					
AAAA9605 standard; DNA: 8993 BP.					
ID	XX	AA	AAA49605:		
DT	XX	AC	19-DEC-2000	(first entry)	
XX	XX				
DE	XX			Mouse villin gene regulatory elements and partial sequence.	
KW	XX			Mouse; villin; intestinal epithelial cell;	
KM	XX			uro-genital tract epithelial cell; tumour; ds.	
OS	XX			Mus sp.	
FH	Key			Location/Qualifiers	
FT	misc_signal	a		/*tag=	
FT		/note= "transcriptional start site"		3442..3487	
FT	exon	b		/*tag=	

FT	/number=1
FT	3488..8981
FT	/*tag=C
FT	/number=1
FT	8982..8995
FT	/*tag=d
FT	/number=2
FT	misc_signal
FT	8992
FT	/*tag=e
FT	/note="translational start site"
XX	
PN	MO200034493-A2.
XX	
PD	15-JUN-2000.
XX	
PF	09-DEC-1999; 99WO-EP09782.
XX	
PR	09-DEC-1998; 98WO-EPO8009.
XX	
PA	(CNRS.) CENT NAT RECH SCI.
PA	(CURT-) INST CURIE.
PI	
P1	Pluto D, Robine S, Jaisser F, Louvard D, Niewoehner J;
DR	
XX	WPI: 2000-423434/36.
XX	
PT	Novel nucleotide sequence derived from mouse villin gene for targeted
PT	expression of transgenes in immature and differentiated epithelial
PT	cells of intestine or urogenital tracts -
XX	
PS	Claim 3; Fig 6; 52pp; English.
XX	
CC	The present sequence consists of the regulatory region and the first
CC	exon and intron of the murine villin gene. This gene is expressed in the
CC	epithelial cells of the intestine and uro-genital tracts. Its promoter
CC	sequence can be used in the targeted expression of exogenous genes in
CC	these places, which may, for example, be useful in the treatment of
CC	tumours.
XX	
SQ	Sequence 8993 BP; 2274 A; 2105 C; 2257 G; 2355 T; 2 other:
	Query Match 96.2%; Score 8651; DB 21; Length 8993:
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 8991; Conservative 0; Mismatches 2; Indels 2; Gaps 2.
OY	1 GATCTGGGCAACCAAGGACTGTGGTCCCAAGCACTGGGGAGTGGAAGGAGGTCGA 60
Dd	
	1 GATCTGGGCAACCAAGGACTGTGGTCCCAAGCACTGGGGAGTGGAAGGAGGTCGA 60
OY	61 GAAGTTTAAGGTCACTCCTTGTTACATAGCAAGAAGTTTCAGGCAGCTTACGTACATAA 120
Dd	61 GAAGTTTAAGGTCACTCCTTGTTACATAGCAAGAAGTTTCAGGCAGCTTACGTACATAA 120
OY	121 CCTTTGTTTTGTTGTTGTTGTTTAAAGCATTAATAATATACCATTAAAGAGAGTTGG 180
Dd	121 CCTTTGTTTTGTTGTTGTTGTTTAAAGCATTAATAATATACCATTAAAGAGAGTTGG 180
OY	181 CAGTGGGGGCGACACCTTTAATGCCAGTATTCAGGAGGCAAGACAGAGATCTCTGT 240
Dd	181 CAGTGGGGGCGACACCTTTAATGCCAGTATTCAGGAGGCAAGACAGAGATCTCTGT 240
OY	241 GAGTTTGAAGTCAGACCTAGTCTGCAAAAGCTAGTTCCAGAGATGGCAAAGGCTACACAGAGA 300
Dd	241 GAGTTTGAAGTCAGACCTAGTCTGCAAAAGCTAGTTCCAGAGATGGCAAAGGCTACACAGAGA 300
OY	301 AACCTTGTCTATAAACCAAGTAGTAGTAGTAGTAATGCCATAGAGAAAATTGGA 360
Dd	301 AACCTTGTCTATAAACCAAGTAGTAGTAGTAGTAATGCCATAGAGAAAATTGGA 360
OY	361 GTTCATTGAGATGAGACATCCTTAATAGATGATTTCTTTGACCCCGTAGTAATGTCA 420
Dd	361 GTTCATTGAGATGAGACATCCTTAATAGATGATTTCTTTGACCCCGTAGTAATGTCA 420



QY 421 TGGGAAAGGGATGGGACTGCTAGATTAAAAAGTGGTGAAGCGATGGCTATTCTCAA 480  
DB 421 TGGGAAAGGGATGGGACTGCTAGATTAAAAAGTGGTGAAGCGATGGCTATTCTCAA 480  
QY 481 TTTTGATTCATATGAAAAAGGCTGTATAGGCCCAAGAGAAAGTGAAGTGGGACTGTGACT 540  
DB 481 TTTTGATTCATATGAAAAAGGCTGTATAGGCCCAAGAGAAAGTGAAGTGGGACTGTGACT 540  
QY 541 GAAGAGGTGACGGCTTTATTAACACTGGCCTTATTAACACTTATTAACACTGGCAGG 600  
DB 541 GAAGAGGTGACGGCTTTATTAACACTGGCCTTATTAACACTTATTAACACTGGCAGG 600  
QY 601 CGTTCAAGTTTGAAGATCACTTTCAACACACAGAAAGAAAGTGGTGGTCTGCTCAGC 660  
DB 601 CGTTCAAGTTTGAAGATCACTTTCAACACACAGAAAGAAAGTGGTGGTCTGCTCAGC 660  
QY 661 GTAGCGACGACTGGCTGACAGAAAGATATTTAGTGAAGAAAGTACCTTCAATATCTTT 720  
DB 661 GTAGCGACGACTGGCTGACAGAAAGATATTTAGTGAAGAAAGTACCTTCAATATCTTT 720  
QY 721 GCACATTATACATACACAGTGTCAAAATGTGCTAACTCCCTAGTCCACAGATGGCTTACA 780  
DB 721 GCACATTATACATACACAGTGTCAAAATGTGCTAACTCCCTAGTCCACAGATGGCTTACA 780  
QY 781 CTCGTTTCTGCTTTCCCATCTGGTTGACATTTGTGCAAAACAGAAATTTAGAAATGTGGT 840  
DB 781 CTCGTTTCTGCTTTCCCATCTGGTTGACATTTGTGCAAAACAGAAATTTAGAAATGTGGT 840  
QY 841 ATTTATTTGTGTGCTGACAGACACATCCAGGGCTTTTCACATTTCAAGGCATATGTTTAC 900  
DB 841 ATTTATTTGTGTGCTGACAGACACATCCAGGGCTTTTCACATTTCAAGGCATATGTTTAC 900  
QY 901 TTAAGTGGGCTACTTCTCCAAAGGTTTGAACCATTGTTTATATTTACTTATTTTGTGT 960  
DB 901 TTAAGTGGGCTACTTCTCCAAAGGTTTGAACCATTGTTTATATTTACTTATTTTGTGT 960  
QY 961 GCATGAGGTAGGCGATGTATACGTATGTATAGAGTCAATGCAATGGCTGCTACCCCTCAA 1020  
DB 961 GCATGAGGTAGGCGATGTATACGTATGTATAGAGTCAATGCAATGGCTGCTACCCCTCAA 1020  
QY 1021 ATCATTTGCAAGTCCCGACAGTGAAGTGAACCGAGGGTGTATATGTTAGTGGAGCTG 1080  
DB 1021 ATCATTTGCAAGTCCCGACAGTGAAGTGAACCGAGGGTGTATATGTTAGTGGAGCTG 1080  
QY 1081 GGAGCCCAAGGCTGGTCTCTGCAAGAGCAGCAGTGGCTTAAACATGGGACACACTCT 1140  
DB 1081 GGAGCCCAAGGCTGGTCTCTGCAAGAGCAGCAGTGGCTTAAACATGGGACACACTCT 1140  
QY 1141 CTAGGCTTAAGGTAAATCTTTAGTTTAAAAATATATATCTCAGCCGGGTGTGGTGC 1200  
DB 1141 CTAGGCTTAAGGTAAATCTTTAGTTTAAAAATATATATCTCAGCCGGGTGTGGTGC 1200  
QY 1201 ACACGGCTTAATCCCGACAGTGAAGGCTGAGGTAGGAATTTATACACACAGGCCAG 1260  
DB 1201 ACACGGCTTAATCCCGACAGTGAAGGCTGAGGTAGGAATTTATACACACAGGCCAG 1260  
QY 1261 CTGGGGTGCAGAGCTTGGCCCTGTCTTTTGTCTTTTCTTTATGCACTGGTGTCTTA 1320  
DB 1261 CTGGGGTGCAGAGCTTGGCCCTGTCTTTTGTCTTTTCTTTATGCACTGGTGTCTTA 1320  
QY 1321 CCTGGCTGTATGCTGCTGCAAGGGTGTCAAGTCCCTGGAGCTGGAATTAAAGACAGTTG 1380  
DB 1321 CCTGGCTGTATGCTGCTGCAAGGGTGTCAAGTCCCTGGAGCTGGAATTAAAGACAGTTG 1380  
QY 1381 TGAATCAGCTGCGCTTACAGATGCTGGAATTTGAACCCAGGTGTCCCTAGAGAGAGGCC 1440  
DB 1381 TGAATCAGCTGCGCTTACAGATGCTGGAATTTGAACCCAGGTGTCCCTAGAGAGAGGCC 1440  
QY 1441 AGTGTCTTAATCTTGAAGCACCCTCCAAACCTCTTTAGAGACTCTTAACCTTTTG 1500  
DB 1441 AGTGTCTTAATCTTGAAGCACCCTCCAAACCTCTTTAGAGACTCTTAACCTTTTG 1500  
QY 1501 TGTAAATGTGGAACTGAAGTGAATCTTGCACTTAACAAAGTGTGTGCTGCTGTAGCATCA 1560

DB 1501 TGTAAATGTGGAACTGAAGTGAATCTTGCACTTAACAAAGTGTGTGCTGCTGTAGCATCA 1560  
QY 1561 CTGAGGCCGCTACCCACACAGCATGTAGTATACATTTTAAGGCAAAACACTTAACAATGACA 1620  
DB 1561 CTGAGGCCGCTACCCACACAGCATGTAGTATACATTTTAAGGCAAAACACTTAACAATGACA 1620  
QY 1621 ATAGTTGATAGAGTTTGAATATAGTCCAGTATTTGTTAGCGTACCTTGTGCTGTC 1680  
DB 1621 ATAGTTGATAGAGTTTGAATATAGTCCAGTATTTGTTAGCGTACCTTGTGCTGTC 1680  
QY 1681 TTAGCATGTGCTGTGAGAAAGATGAAAAATGAAGACTTGAATCTGAGAACCCACA 1740  
DB 1681 TTAGCATGTGCTGTGAGAAAGATGAAAAATGAAGACTTGAATCTGAGAACCCACA 1740  
QY 1741 GAGGCAAGGAGAGAACCCCTCCGAAAAGTTGTTCTGAGCTTCACATACACTTACAT 1800  
DB 1741 GAGGCAAGGAGAGAACCCCTCCGAAAAGTTGTTCTGAGCTTCACATACACTTACAT 1800  
QY 1801 AATAGTTTACATGATATATATATATATATTTAGTAATTTCTTTAAAGGTATATGTTGGAGGA 1860  
DB 1801 AATAGTTTACATGATATATATATATATTTAGTAATTTCTTTAAAGGTATATGTTGGAGGA 1860  
QY 1861 GAGATGGCTCAGCTTCCAGAGGACCTTGTCTTTGACAGAGGACCTAGATTGCTGCC 1920  
DB 1861 GAGATGGCTCAGCTTCCAGAGGACCTTGTCTTTGACAGAGGACCTAGATTGCTGCC 1920  
QY 1921 AGGACTCATATGTTGGCTCAGACCCATCTGTAAATCCAGTCCAGAGAGGTTCCACACCT 1980  
DB 1921 AGGACTCATATGTTGGCTCAGACCCATCTGTAAATCCAGTCCAGAGAGGTTCCACACCT 1980  
QY 1981 CTTCGCGCTCCACAGGACACACATATATATATATATATATATATATATATATATATATAT 2040  
DB 1981 CTTCGCGCTCCACAGGACACACATATATATATATATATATATATATATATATATATATAT 2040  
QY 2041 CATACACATTAATTAATTAAGAACTTAAGAAAGTGTGATGTTGTTGAACATTTGCT 2100  
DB 2041 CATACACATTAATTAATTAAGAACTTAAGAAAGTGTGATGTTGTTGAACATTTGCT 2100  
QY 2101 TACACATGCTGATTAAGAGATGTACAGGACACACTGAAAGAGGATCTGGGGCTGGAG 2160  
DB 2101 TACACATGCTGATTAAGAGATGTACAGGACACACTGAAAGAGGATCTGGGGCTGGAG 2160  
QY 2161 AGATGGCTCAGCGGTTAAGAGCACTGACTCTTCCGAAGGAAGTCTGATTTCAAT 2220  
DB 2161 AGATGGCTCAGCGGTTAAGAGCACTGACTCTTCCGAAGGAAGTCTGATTTCAAT 2220  
QY 2221 CCTAGCAACCATGTTGGCTCACAACCATCATTAATGAGATGTGACACCTTCTGT 2280  
DB 2221 CCTAGCAACCATGTTGGCTCACAACCATCATTAATGAGATGTGACACCTTCTGT 2280  
QY 2281 GCATCTGAAGACGCTGCAAGAGCTACAGTACTTGTAGATTAATTAATTAATCTTTT 2340  
DB 2281 GCATCTGAAGACGCTGCAAGAGCTACAGTACTTGTAGATTAATTAATTAATCTTTT 2340  
QY 2341 TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTAAGCAGTACTCAGC 2400  
DB 2341 TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTAAGCAGTACTCAGC 2400  
QY 2401 GGTGATTAATCTATCTGAGAGTTTTCCTTTCCGCTTGGCTTGCACATGGGTGACAGAGC 2460  
DB 2401 GGTGATTAATCTATCTGAGAGTTTTCCTTTCCGCTTGGCTTGCACATGGGTGACAGAGC 2460  
QY 2461 CCCCTTTTCAATCACAAGAAAGGCTCTCATATTTTGTGACAAAAGCAACCCCTGACG 2520  
DB 2461 CCCCTTTTCAATCACAAGAAAGGCTCTCATATTTTGTGACAAAAGCAACCCCTGACG 2520  
QY 2521 ATGTTACTGTCTCTTCTACTATGTAGCAGCGCAGCGCGCGCAGCACAACACACAC 2580  
DB 2521 ATGTTACTGTCTCTTCTACTATGTAGCAGCGCAGCGCGCGCAGCACAACACACAC 2580  
QY 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAAGTCA 2640

Db 2581 ACACACACACACACACACACACATTCAGTCTCCAGAGCTTGTGGGAGGTCA 2640  
QY 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTTAAAGAGACAGATTC 2700  
Db 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTTAAAGAGACAGATTC 2700  
QY 2701 AAGGTGGCAGAGATCTACAGGGGGCAGAGCCAGGAGGGGGAACAGAGCCATGGTTCC 2760  
Db 2701 AAGGTGGCAGAGATCTACAGGGGGCAGAGCCAGGAGGGGGAACAGAGCCATGGTTCC 2760  
QY 2761 AGAGACCTACAGAGAGGGGCAAGCAAGTCCAGTCCAGGGGGCAGGAGGGGAGG 2820  
Db 2761 AGAGACCTACAGAGAGGGGCAAGCAAGTCCAGTCCAGGGGGCAGGAGGGGAGG 2820  
QY 2821 CCCTTGTTCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2880  
Db 2821 CCCTTGTTCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2880  
QY 2881 GCTCATTAAGTGAAGTCTGTGGCTCACTCAGAAAGAGAGAGAGAGAGAGGAGGAGG 2940  
Db 2881 GCTCATTAAGTGAAGTCTGTGGCTCACTCAGAAAGAGAGAGAGAGAGGAGGAGG 2940  
QY 2941 GCCCAGCTGAGGAGGAGTCTGTAGTAGAGAGATCGCAAGGGGGTCCAGAGAGCCAC 3000  
Db 2941 GCCCAGCTGAGGAGGAGTCTGTAGTAGAGAGATCGCAAGGGGGTCCAGAGAGCCAC 3000  
QY 3001 CTGTCTGTCCCAAGGAGACCCCAAGTGAAGTGAAGTCTGTGGCTGTGTAGTTCAGCTA 3060  
Db 3001 CTGTCTGTCCCAAGGAGACCCCAAGTGAAGTGAAGTCTGTGGCTGTGTAGTTCAGCTA 3060  
QY 3061 CAAAGACCCAGAGAGTCTACTTCATCCCATTCAGTGGCCCTGGCCGCCACACCCCA 3120  
Db 3061 CAAAGACCCAGAGAGTCTACTTCATCCCATTCAGTGGCCCTGGCCGCCACACCCCA 3120  
QY 3121 CCCCCAGTCCCGTCCACTCTGTAGGGGTGAGGGGTGGCAGAGCCCTGGGGGGGTTC 3180  
Db 3121 CCCCCAGTCCCGTCCACTCTGTAGGGGTGAGGGGTGGCAGAGCCCTGGGGGGGTTC 3180  
QY 3181 CTACCTGAGGTAGAGCCAGAGTCTAGCCGGAAGTGCACCCCATCCTGAAGTGCAGA 3240  
Db 3181 CTACCTGAGGTAGAGCCAGAGTCTAGCCGGAAGTGCACCCCATCCTGAAGTGCAGA 3240  
QY 3241 GCCAAGGGGGGGGACACAGGAGCTCAGAGCTGTAGGGCTGTGGCTCTAGGTTC 3300  
Db 3241 GCCAAGGGGGGGGACACAGGAGCTCAGAGCTGTAGGGCTGTGGCTCTAGGTTC 3300  
QY 3301 AGGAGCTGGGACCTACTTCCCAACCCCATTCATCTCTGTGGGGCCCTATCTCC 3360  
Db 3301 AGGAGCTGGGACCTACTTCCCAACCCCATTCATCTCTGTGGGGCCCTATCTCC 3360  
QY 3361 CTATATGTGTAAGAGATCTCTGTGGGGGGGGGGGGTGTGTGAGACAAAGTCTTCG 3420  
Db 3361 CTATATGTGTAAGAGATCTCTGTGGGGGGGGGGGGTGTGTGAGACAAAGTCTTCG 3420  
QY 3421 GTTCCTGTGACAGCCAGTCCCAACTTCTGAAGTCTCCAGAGGTGGGCTCTTC 3480  
Db 3421 GTTCCTGTGACAGCCAGTCCCAACTTCTGAAGTCTCCAGAGGTGGGCTCTTC 3480  
QY 3481 CAGACAGGTAAAGGCAATTGGGTGGGGACACATGTTGACACAGTGTGGAGGGAGACG 3540  
Db 3481 CAGACAGGTAAAGGCAATTGGGTGGGGACACATGTTGACACAGTGTGGAGGGAGACG 3540  
QY 3541 GGTCCTGTCTCTGTGGACAGCTGTCTTCTGTAGACCTTGGTATAGTTGGGGG 3600  
Db 3541 GGTCCTGTCTCTGTGGACAGCTGTCTTCTGTAGACCTTGGTATAGTTGGGGG 3600  
QY 3601 TGAGGTAGGTGCTGTGAAGTCTGAAGAGCAAGAGCAAGGAGGAGGAGGAGGAGG 3660  
Db 3601 TGAGGTAGGTGCTGTGAAGTCTGAAGAGCAAGAGCAAGGAGGAGGAGGAGGAGG 3660  
QY 3661 TCAATGAAGAGATTCACAGAGACCCCTTCTGTAACTACCTTCATCTGTGTAG 3720  
Db 3661 TCAATGAAGAGATTCACAGAGACCCCTTCTGTAACTACCTTCATCTGTGTAG 3720

QY 3721 ATTCCTGTGGACCAAGGTGGCTCTGTGGACTCAGATTTCTACAAATTAATCAGAGAGT 3780  
Db 3721 ATTCCTGTGGACCAAGGTGGCTCTGTGGACTCAGATTTCTACAAATTAATCAGAGAGT 3780  
QY 3781 CCTGAGACTTGAGCTCCGTGCTGTATTTACTACTTCTCTGTGGCTGTCTATTTCTGT 3840  
Db 3781 CCTGAGACTTGAGCTCCGTGCTGTATTTACTACTTCTCTGTGGCTGTCTATTTCTGT 3840  
QY 3841 TCATGTCTTACACATCTGAATGTTTCTTGTGTGACACATTTCCCTGACACTCTGTGGA 3900  
Db 3841 TCATGTCTTACACATCTGAATGTTTCTTGTGTGACACATTTCCCTGACACTCTGTGGA 3900  
QY 3901 GGTCGTATCTTGGACACATGTATCTGTGGATGTAACTGACAGCCACAGAGAGAGGGG 3960  
Db 3901 GGTCGTATCTTGGACACATGTATCTGTGGATGTAACTGACAGCCACAGAGAGAGGGG 3960  
QY 3961 AGAGTCAGAGAGTGTGTCTAGGGCCCTTTAGGCTGTGACATTCACCCCTTCTGAAGAT 4020  
Db 3961 AGAGTCAGAGAGTGTGTCTAGGGCCCTTTAGGCTGTGACATTCACCCCTTCTGAAGAT 4020  
QY 4021 GGCCCTTCATTTTTCGTTACATGATCTATTTATATCAGAGTGGGACAGTGAAGGCA 4080  
Db 4021 GGCCCTTCATTTTTCGTTACATGATCTATTTATATCAGAGTGGGACAGTGAAGGCA 4080  
QY 4081 AACCTGCCAGAGTGTGGAGCTCACTCAGACCAAGGTTATCTGCTAGAAATCCCTCG 4140  
Db 4081 AACCTGCCAGAGTGTGGAGCTCACTCAGACCAAGGTTATCTGCTAGAAATCCCTCG 4140  
QY 4141 TCACCTGAGGTGGGAGAAATCGCTCTGGGGGCTTCAGAGTCTTGGTTACAGAGAGGT 4200  
Db 4141 TCACCTGAGGTGGGAGAAATCGCTCTGGGGGCTTCAGAGTCTTGGTTACAGAGAGGT 4200  
QY 4201 ATCCTTTGTATAGGGCATGACCTAGTCTATGGTGTACTACATTCCTGACGTTAAAG 4260  
Db 4201 ATCCTTTGTATAGGGCATGACCTAGTCTATGGTGTACTACATTCCTGACGTTAAAG 4260  
QY 4261 CTGGNACTAAACCAGGAGGAGCCGCGAGATTTCTTACAGTTGTACCCCAAGAACAA 4320  
Db 4261 CTGGNACTAAACCAGGAGGAGCCGCGAGATTTCTTACAGTTGTACCCCAAGAACAA 4320  
QY 4321 AGACAGTGAATGCAAGGATGAGTGTGGGGGAGAAACAACTTAACCCGCCCAAG 4380  
Db 4321 AGACAGTGAATGCAAGGATGAGTGTGGGGGAGAAACAACTTAACCCGCCCAAG 4380  
QY 4381 GCCCAGAGTTCGTTCCCTAGTTCACAATGCCAGTATGAGTGTAGCTACTATGGGCTG 4440  
Db 4381 GCCCAGAGTTCGTTCCCTAGTTCACAATGCCAGTATGAGTGTAGCTACTATGGGCTG 4440  
QY 4441 TGAGTTGTGATACACAGATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4500  
Db 4441 TGAGTTGTGATACACAGATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4500  
QY 4501 TTGGAGAGCTGAAGAGAGGATGCTATATGTTTGAAGCCAGCTGAGCTATGAGAGGA 4560  
Db 4501 TTGGAGAGCTGAAGAGAGGATGCTATATGTTTGAAGCCAGCTGAGCTATGAGAGGA 4560  
QY 4561 GACTTTGTCTTAAAGAAAAATGAAGCCAGCAGTGTGGACACCCCTTTATATCCCA 4620  
Db 4561 GACTTTGTCTTAAAGAAAAATGAAGCCAGCAGTGTGGACACCCCTTTATATCCCA 4620  
QY 4621 GCACCTTGGGAGGAGAGGAGGATTTCTGTAGTTCAGAGCCAGCCGTGTATAGAGT 4680  
Db 4621 GCACCTTGGGAGGAGAGGAGGATTTCTGTAGTTCAGAGCCAGCCGTGTATAGAGT 4680  
QY 4681 GAGTTCAGAGAGGAGGAGGATTCAGAGAGAAACCTTTTGAAGAAAAACAGAAAAACA 4740  
Db 4681 GAGTTCAGAGAGGAGGAGGATTCAGAGAGAAACCTTTTGAAGAAAAACAGAAAAACA 4740  
QY 4741 AACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACCTCTCATCTCTCTC 4800  
Db 4741 AACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACCTCTCATCTCTCTC 4800

QY	4801	TAGCCTGTGTCTGTCTAGTGGTAGAGCTTTGGGACCTTCAACATTATATATAATAGGCC	4860
Db	4801	TAGCCTGTGTCTGTCTAGTGGTAGAGCTTTGGGACCTTCAACATTATATATAATAGGCC	4860
QY	4861	TTTTATACACGCTAGAGACGAGAAAGTTTCACTGTCGGGACACAGTGGGACCTGGA	4920
Db	4861	TTTTATACACGCTAGAGACGAGAAAGTTTCACTGTCGGGACACAGTGGGACCTGGA	4920
QY	4921	AAGTACTCTTCCAGCCCAAAAATTCGGGAAAGCTTCCGGAAGAAAGTGTGCCGAT	4980
Db	4921	AAGTACTCTTCCAGCCCAAAAATTCGGGAAAGCTTCCGGAAGAAAGTGTGCCGAT	4980
QY	4981	CAGACTACTGTTCTTGAAGGCAAGAGAGGCTTGGAAATGTTGGTGCACACAGTT	5040
Db	4981	CAGACTACTGTTCTTGAAGGCAAGAGAGGCTTGGAAATGTTGGTGCACACAGTT	5040
QY	5041	GGAAACAGAAAGCAGAGAGGGGAGGCATCCAAAGATTCGAACATGTAAGTCACTTTGGT	5100
Db	5041	GGAAACAGAAAGCAGAGAGGGGAGGCATCCAAAGATTCGAACATGTAAGTCACTTTGGT	5100
QY	5101	TCTCTGGGTGACAAAGTGTCCCCCAGAGGATAGGGCTGTGAAGAAAGGGACACAGGGTGAACC	5160
Db	5101	TCTCTGGGTGACAAAGTGTCCCCCAGAGGATAGGGCTGTGAAGAAAGGGACACAGGGTGAACC	5160
QY	5161	AATGACTTCAAGTTGAGGGGACACATCCAGCCAGGGTCTTGCTGGCAAGCTAAAGAAATG	5220
Db	5161	AATGACTTCAAGTTGAGGGGACACATCCAGCCAGGGTCTTGCTGGCAAGCTAAAGAAATG	5220
QY	5221	AGAGCCCTTAACCCCTCCCTGAAGTTTAGGGGACACAGAGAGCTGAGAGAAATCCTCTTA	5280
Db	5221	AGAGCCCTTAACCCCTCCCTGAAGTTTAGGGGACACAGAGAGCTGAGAGAAATCCTCTTA	5280
QY	5281	GGGTGAAGAGAGAGTATGCTCTGACCAATGGCTGAGAGCAAGACATTTGGACAG	5340
Db	5281	GGGTGAAGAGAGAGTATGCTCTGACCAATGGCTGAGAGCAAGACATTTGGACAG	5340
QY	5341	TTACCCCTCAGAACCCAGCCATCCCTCTGGGCTCTAAGAGGCTGGGCCCTTCTGT	5400
Db	5341	TTACCCCTCAGAACCCAGCCATCCCTCTGGGCTCTAAGAGGCTGGGCCCTTCTGT	5400
QY	5401	AAGATCTTACTTTCTTCACAGAGAGGACACAAAGCTTTGCCCTCCCTGTTGTCTAA	5460
Db	5401	AAGATCTTACTTTCTTCACAGAGAGGACACAAAGCTTTGCCCTCCCTGTTGTCTAA	5460
QY	5461	TAAACACCCCTGTGTGAACATTAAGTTATTTACTGTCACTTTGGTCTCAGAGACATCCA	5520
Db	5461	TAAACACCCCTGTGTGAACATTAAGTTATTTACTGTGTCACTTTGGTCTCAGAGACATCCA	5520
QY	5521	TCTGTAGACCTGTGCTCTTAACCTCACCAAGATATGGCCCATTTCTCCACCCAGAGAG	5580
Db	5521	TCTGTGTAGACCTGTGCTCTTAACCTCACCAAGATATGGCCCATTTCTCCACCCAGAGAG	5580
QY	5581	TGCAGAAAGAGCCTTAGAGAAAGGGTAAACGTAACAAAGATGGCCAGATATAAACAAA	5640
Db	5581	TGCAGAAAGAGCCTTAGAGAAAGGGTAAACGTAACAAAGATGGCCAGATATAAACAAA	5640
QY	5641	ACTACTATCCCTTTGTACCCAAATTTGTTTTCTGTAAACAGAGAGGGGGTGTGTAGTGTAT	5700
Db	5641	ACTACTATCCCTTTGTACCCAAATTTGTTTTCTGTAAACAGAGAGGGGGTGTGTAGTGTAT	5700
QY	5701	GTTGT	5760
Db	5701	GTTGT	5760
QY	5761	CTTGGGGACCTTTTCATCTAAAGAAATATCGATAATTTGGCCCATGTCCAAGGGGTAT	5820
Db	5761	CTTGGGGACCTTTTCATCTAAAGAAATATCGATAATTTGGCCCATGTCCAAGGGGTAT	5820
QY	5821	TGGGAGAGTACAGGCTTCTGCAAAACACAGTAAGCTGCCCAAGATGGAATTTGGTGGCTGAA	5880
Db	5821	TGGGAGAGTACAGGCTTCTGCAAAACACAGTAAGCTGCCCAAGATGGAATTTGGTGGCTGAA	5880
QY	5881	TCACCAAGGGGACAGGCTGATCAGATGGACAGAACATCCAAAGATTAAGCCACCCCTGTGGG	5940

[illegible]

D 6960 CACAGATGACCTTTAACTCCTGCTCTTCCAGCCTCCACCCGAGTGTAGTTTACAGGT 7019  
Q 7021 GTTCAACTGGTGAATGCTTTTAAATCCAGCACTCTGTGGGGGGGGGGAGCGGATC 7080  
D 7020 GTTCAACTGGTGAATGCTTTTAAATCCAGCACTCTGTGGGGGGGGGGAGCGGATC 7079  
Q 7081 CTTGAGTTGGAGGCGCACTTTGGTCTACAGAGTTTACAGATACCTGGGGCTATACAGGAA 7140  
D 7080 CTTGAGTTGGAGGCGCACTTTGGTCTACAGAGTTTACAGATACCTGGGGCTATACAGGAA 7139  
Q 7141 ACCCTATCCCAACAAACAACAACAACAACAAAATAATCTGTGCAATATCAGAGAA 7200  
D 7140 ACCCTATCCCAACAAACAACAACAACAACAAAATAATCTGTGCAATATCAGAGAA 7199  
Q 7201 TTAGAGATATTTAGTAGGTAAGTAGGGCTGTGAGGAGATCAGCTTTCTTTTGTATT 7260  
D 7200 TTAGAGATATTTAGTAGGTAAGTAGGGCTGTGAGGAGATCAGCTTTCTTTTGTATT 7259  
Q 7261 ATATAGTAAAGTACTACAGAGATGCAATATCTATCTATCTATCTATCTATCTATC 7320  
D 7260 ATATAGTAAAGTACTACAGAGATGCAATATCTATCTATCTATCTATCTATCTATC 7319  
Q 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7380  
D 7320 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7379  
Q 7381 CTGCTTTGACTCTGAATGCTCTAATTTCTGGGTCACCTCTACCCCTAGTGTGGATT 7440  
D 7380 CTGCTTTGACTCTGAATGCTCTAATTTCTGGGTCACCTCTACCCCTAGTGTGGATT 7439  
Q 7441 ACCAACCACAGACATTTATTTATTTTGTATTTATTTATTTATTTATTTATTTATTTAT 7500  
D 7440 ACCAACCACAGACATTTATTTATTTTGTATTTATTTATTTATTTATTTATTTATTTAT 7499  
Q 7501 GGAATCAGGGTCTGTGATGCTAAGCAAGCTCTGTCCACAGAGCTGACGCTCAATCC 7560  
D 7500 GGAATCAGGGTCTGTGATGCTAAGCAAGCTCTGTCCACAGAGCTGACGCTCAATCC 7559  
Q 7561 CCATTTTGTTCAGTACTGTGACAGTTGTGATATTCGACAGGCTATGACTGTCTC 7620  
D 7560 CCATTTTGTTCAGTACTGTGACAGTTGTGATATTCGACAGGCTATGACTGTCTC 7619  
Q 7621 CACCTCCAGTTCCAGACTTTTGTGATCCCAAGTGGGGGGAACCTGTGTGTCACA 7680  
D 7620 CACCTCCAGTTCCAGACTTTTGTGATCCCAAGTGGGGGGAACCTGTGTGTCACA 7679  
Q 7681 GTGCGCTGTCCCTGTGCTTACAGACTATATTTGCTGTGCAAGCTTCAATGAATG 7740  
D 7680 GTGCGCTGTCCCTGTGCTTACAGACTATATTTGCTGTGCAAGCTTCAATGAATG 7739  
Q 7741 GGATGCGTCTCTGTATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7800  
D 7740 GGATGCGTCTCTGTATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7799  
Q 7801 GGCATGTGTACATGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7860  
D 7800 GGCATGTGTACATGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7859  
Q 7861 TGTGTGATTAACCACTTTCTGTTTCAATTTACTGATGGAATTTGTGGCCCAACCCAC 7920  
D 7860 TGTGTGATTAACCACTTTCTGTTTCAATTTACTGATGGAATTTGTGGCCCAACCCAC 7919  
Q 7921 CTTTTTTTTTTTTTATTTGAGACAAGCTTTTGTGTGTAATCTTGCAATCTTGCTGTCT 7980  
D 7920 CTTTTTTTTTTTTTATTTGAGACAAGCTTTTGTGTGTAATCTTGCAATCTTGCTGTCT 7979  
Q 7981 GAGCTCACTCTGTAGACAGAGCTGTGAGCTGTCTCCACTTTTGACACTCCGTGAAC 8040  
D 7980 GAGCTCACTCTGTAGACAGAGCTGTGAGCTGTCTCCACTTTTGACACTCCGTGAAC 8039  
Q 8041 AGAGTACCAATGAATCTCAAGACAAATTTCTGTTTGGTTTGTATTTTCAATTTGTGTG 8100  
D 8040 AGAGTACCAATGAATCTCAAGACAA-TTCTGTGTTTGTGTTTGTATTTTACATTTGTGTG 8098

Q 8101 TGTATGCGTGTATATGTGATGTTTGTGTCTTCAAGTGTCTACATGTGTGTACCTGTGTG 8160  
D 8099 TGTATGCGTGTATATGTGATGTTTGTGTCTTCAAGTGTCTACATGTGTGTACCTGTGTG 8158  
Q 8161 TGGACAGACAAACAACCGATGTGCCATTTCTCTAGATACATACGATCTTTGTATATGTA 8220  
D 8159 TGGACAGACAAACAACCGATGTGCCATTTCTCTAGATACATACGATCTTTGTATATGTA 8218  
Q 8221 TGTATATGTATGTATTTATTTAGTGTGCCAAGTATGAGGATATTTTGTGAGTTTTCAC 8280  
D 8219 TGTATATGTATGTATTTATTTAGTGTGCCAAGTATGAGGATATTTTGTGAGTTTTCAC 8278  
Q 8281 CTTCCCTTGTGGGCTCTCCGCAATTAACCTAGCTCTCTGGGCTATGAGCAATGCCCTTCA 8340  
D 8279 CTTCCCTTGTGGGCTCTCCGCAATTAACCTAGCTCTCTGGGCTATGAGCAATGCCCTTCA 8338  
Q 8341 CTGATATGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8400  
D 8339 CTGATATGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8398  
Q 8401 CGCAGTCACTGGCTTAAGCTTCAACAGTATCAAGTATGAGTATGAGTATGAGTATGAGTAT 8460  
D 8399 CGCAGTCACTGGCTTAAGCTTCAACAGTATCAAGTATGAGTATGAGTATGAGTATGAGTAT 8458  
Q 8461 GGATATCTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8520  
D 8459 GGATATCTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8518  
Q 8521 ATTTTAAACCGAATCTGAGGATGAGACAGGCTCTTACCAATGAGGAGGCTCTTTTGT 8580  
D 8519 ATTTTAAACCGAATCTGAGGATGAGACAGGCTCTTACCAATGAGGAGGCTCTTTTGT 8578  
Q 8581 GTTGTGTTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8640  
D 8579 GTTGTGTTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8638  
Q 8641 ATTAACATCTGTCTCAAAAAGCTATATGAGTATGAGGAGTGTGAGGCTTAAAGAGGCTT 8700  
D 8639 ATTAACATCTGTCTCAAAAAGCTATATGAGTATGAGGAGTGTGAGGCTTAAAGAGGCTT 8698  
Q 8701 AAGCGGCTGTGATGAGCAGACAGATAGGCTGCTATATGAGCAAGCTTGTTCAAA 8760  
D 8699 AAGCGGCTGTGATGAGCAGACAGATAGGCTGCTATATGAGCAAGCTTGTTCAAA 8758  
Q 8761 ACATGAGGAGGAGGATATGTTTATGCTGTGGCTGTGTAAAGCAAGCTTAAAGAGGCCAA 8820  
D 8759 ACATGAGGAGGAGGATATGTTTATGCTGTGGCTGTGTAAAGCAAGCTTAAAGAGGCCAA 8818  
Q 8821 TGTAGACATTTGACTAAGAAAGATCATCATCAAAAGCCGGGTGGGCAAGGTTGG 8880  
D 8819 TGTAGACATTTGACTAAGAAAGATCATCATCAAAAGCCGGGTGGGCAAGGTTGG 8878  
Q 8881 ACTACAGTGTGCAAGACCCCATATGGAAGCAGTTTCCCTTCTCTGAGGCTCTCAAGC 8940  
D 8879 ACTACAGTGTGCAAGACCCCATATGGAAGCAGTTTCCCTTCTCTGAGGCTCTCAAGC 8938  
Q 8941 CTGGCTGAGCGGCACTGTCTCAACATGCTTCTCTAGGCTGTCTCAACATG 8995  
D 8939 CTGGCTGAGCGGCACTGTCTCAACATGCTTCTCTAGGCTGTCTCAACATG 8993

RESULT 3  
AAS19044  
ID AAS19044 standard; cDNA: 2628 BP.  
XX  
AC AAS19044;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE Mouse MU-1 haematopoietin receptor superfamily chain cDNA sequence.  
XX  
KW Mouse; MU-1; haematopoietin receptor superfamily chain; thyroiditis;



[illegible][illegible]



XX	25-MAR-2003	(updated)
DT	14-JUL-1994	(first entry)
XX		
DE	Grapevine ribosomal clone INVGT19C.	
XX		
KW	Ribosome; grapevine; vitis; IGS region; rDNA; polymorphism;	
KM	grape cultivar; probe; primer; detection; ds.	
XX		
OS	Vitis vinifera.	
XX		
FH	Key	Location/Qualifiers
FT	primer_bind	76..100
FT	primer_bind	/*tag=a
FT	primer_bind	complement (295..317)
XX		/*tag=b
PN	MO9401580-A1.	
XX		
PD	20-JAN-1994.	
XX		
PF	30-JUN-1993;	93WO-AU00320.
XX		
PR	03-JUL-1992;	92AU-0003330.
XX		
PA	(CSTR ) COMMONWEALTH SCT & IND RES ORG.	
XX		
PI	Scott NS, Thomas MR;	
XX		
DR	WPI: 1994-035083/04.	
XX		
PT	Novel ribosome DNA probe sequences - for the accurate identification	
PT	of grape cultivars	
XX		
PS	Claim 26; Fig 5p; 55pp; English.	
XX		
CC	The sequences given in AA055231-50 are derived from ribosomes of the	
CC	grapevine genus Vitis. These sequences represent the IGS region of	
CC	the ribosomal (r)DNA repeat and contain polymorphisms. These	
CC	polymorphisms may be used in a method for the identification of	
CC	different grape cultivars. These clones contain simple repeat	
CC	sequences and were identified in a genomic library of grapevine DNA	
CC	using simple di-, tri- or tetra- nucleotide repeats such as (AT)8,	
CC	(GT)10, (CCT)10 and such like as probes.	
CC	(updated on 25-MAR-2003 to correct PN field.)	
XX		
SO	Sequence 382 BP; 122 A; 28 C; 95 G; 137 T; 0 other;	
	Query Match	0.7%; Score 63; DB 15; Length 382;
	Best Local Similarity	100.0%; Pred. No. 3.1e-16;
	Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	5698 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	575
DB	140 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	199
QY	5758 TGT 5760	
DB	200 TGT 202	
RESULT 10		
AAA74316/C		
ID	AAA74316 standard; DNA; 382 BP.	
XX		
AC	AAA74316;	
XX		
DT	29-NOV-2000 (first entry)	
XX		
DE	loblolly pine SSR locus RIPPT1125.	
XX		
KW	loblolly pine; Simple Sequence Repeat; SSR; microsatellite DNA repeat;	
KM	genetic marker; mapping; inheritance study; population genetics study;	

[illegible]











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:01 ; Search time 31311 Seconds

(without alignments)  
11752.486 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctggtgaccacaagaca.....ctctagctcgtccaccatg 8995

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Genembl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_com: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_com: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pin: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_hgo\_hum: \*  
40: em\_hgo\_mus: \*  
41: em\_hgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query				ID	Description
	Score	Match	Length	DB		
1	8995	100.0	8995	6	AX027824	AX027824 Sequence
2	4465	49.6	192060	2	AC098570	AC098570 Mus muscu
3	2086	23.2	170074	2	AC132885	AC132885 Mus muscu
4	180	2.0	779	11	BV067438	BV067438 S212P6046
5	96	1.1	149109	2	AC101018	AC101018 Rattus no
6	96	1.1	193472	2	AC115173	AC115173 Rattus no
7	82	0.9	145489	2	AC136730	AC136730 Mus muscu
8	82	0.9	154704	2	AC115850	AC115850 Mus muscu
9	82	0.9	173892	2	AC120393	AC120393 Mus muscu
10	82	0.9	189241	2	AC139041	AC139041 Mus muscu
11	82	0.9	190776	2	AC131335	AC131335 Mus muscu
12	82	0.9	203855	10	AC122900	AC122900 Mus muscu
13	82	0.9	222185	2	AC127697	AC127697 Mus muscu
14	82	0.9	226909	2	AC079490	AC079490 Mus muscu
15	82	0.9	232568	2	AC141471	AC141471 Mus muscu
16	82	0.9	253783	2	AC093351	AC093351 Mus muscu
17	81	0.9	181600	2	AC119189	AC119189 Mus muscu
18	81	0.9	216199	2	AC114583	AC114583 Mus muscu
19	81	0.9	247680	2	AC139943	AC139943 Mus muscu
20	80	0.9	192182	2	AC118017	AC118017 Mus muscu
21	80	0.9	235302	2	AC073784	AC073784 Mus muscu
22	80	0.9	316822	2	AC118016	AC118016 Mus muscu
23	79	0.9	57957	10	BX294655	BX294655 Mouse DNA
24	79	0.9	182061	2	AC102217	AC102217 Mus muscu
25	79	0.9	202855	10	AL589737	AL589737 Mouse DNA
26	79	0.9	223290	2	AC140409	AC140409 Mus muscu
27	78	0.9	153332	2	AC117755	AC117755 Mus muscu
28	78	0.9	204068	2	AC103395	AC103395 Mus muscu
29	78	0.9	292653	2	AC102669	AC102669 Mus muscu
30	78	0.9	316888	2	AC102670	AC102670 Mus muscu
31	77	0.9	172956	2	AC122433	AC122433 Mus muscu
32	77	0.9	182777	2	AC110034	AC110034 Mus muscu
33	76	0.8	116898	10	AL731687	AL731687 Mouse DNA
34	76	0.8	133296	10	AL935323	AL935323 Mouse DNA
35	76	0.8	178032	2	AC087869	AC087869 Mus muscu
36	76	0.8	195320	2	AC132412	AC132412 Mus muscu
37	76	0.8	199019	10	AL928812	AL928812 Mouse DNA
38	76	0.8	201602	2	AC074146	AC074146 Mus muscu
39	76	0.8	203083	2	AC069459	AC069459 Mus muscu
40	75	0.8	824	11	BV025926	BV025926 S212P6802
41	75	0.8	71848	2	AC114623_3	Continuation (4 of
42	75	0.8	131685	2	AC142273	AC142273 Mus muscu
43	75	0.8	149992	2	AC112940	AC112940 Mus muscu
44	75	0.8	168889	10	AL731779	AL731779 Mouse DNA
45	75	0.8	186858	10	AC130210	AC130210 Mus muscu

#### ALIGNMENTS

RESULT 1  
AX027824  
LOCUS AX027824 8995 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent WO0034492.  
ACCESSION AX027824  
VERSION AX027824.1 GI:10188668  
KEYWORDS  
SOURCE  
ORGANISM Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 Robine,S., Louvard,D., Pinto,D. and Jaisser,F.  
AUTHORS Regulatory sequences of the mouse villin gene - use in transgenesis  
TITLE Patent: WO 0034492-A 1 15-JUN-2000;  
JOURNAL

Pred. No. is the number of results predicted by chance to have a

ROBINE SYLVIE (FR) ; INST CURIE (FR) ; LOUYARD DANIEL (FR) ; PINTO DANIEL (FR) ; CENTRE NAT RECH SCIENT (FR) ; JAISSER FREDERIC (FR)

## FEATURES

Location/Qualifiers

1..8995  
/organism="Mus sp.;"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10095"

## exon

3443..3487  
/note="exon 1"

## Intron

3489..8981

## BASE COUNT

2275 a 2105 c 2258 g 2357 t

Query Match 100.0%; Score 8995; DB 6; Length 8995;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GATCTGGGACACAAAGGACACTGTGTCGCCAGCACTGGGGAGGGAGGAGGAGGCA 60  
1 GATCTGGGACACAAAGGACACTGTGTCGCCAGCACTGGGGAGGGAGGAGGAGGCA 60  
61 GAAGTTTAAAGGTCACTCTGTTACATAGCAAGGTTTCAGCCAGCTTACGATGAAA 120  
61 GAAGTTTAAAGGTCACTCTGTTACATAGCAAGGTTTCAGCCAGCTTACGATGAAA 120  
121 CCTTTGTTTGTGTTGTTGTTGTTTAAAGCATTAAATTAATTAATTAATTAATTA 180  
121 CCTTTGTTTGTGTTGTTGTTGTTTAAAGCATTAAATTAATTAATTAATTAATTA 180  
181 CAGTGGGGACAGACCTTTAATTCAGATATTCAGAGGGAGGAGGAGGAGGAGGAGG 240  
181 CAGTGGGGACAGACCTTTAATTCAGATATTCAGAGGGAGGAGGAGGAGGAGGAGG 240  
241 GAGTTGCAAGTCAAGCTAGTCTGCAAAAGCTAGTTCAGAGGATGGCAAGGGCTACACAGAGA 300  
241 GAGTTGCAAGTCAAGCTAGTCTGCAAAAGCTAGTTCAGAGGATGGCAAGGGCTACACAGAGA 300  
241 GAGTTGCAAGTCAAGCTAGTCTGCAAAAGCTAGTTCAGAGGATGGCAAGGGCTACACAGAGA 300  
241 GAGTTGCAAGTCAAGCTAGTCTGCAAAAGCTAGTTCAGAGGATGGCAAGGGCTACACAGAGA 300  
301 AACCTTGTCTATTAACCAAGTATGTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 360  
301 AACCTTGTCTATTAACCAAGTATGTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 360  
301 AACCTTGTCTATTAACCAAGTATGTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 360  
301 AACCTTGTCTATTAACCAAGTATGTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 360  
361 GTCCATTGAGGATGACCATCTTAATTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 420  
361 GTCCATTGAGGATGACCATCTTAATTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 420  
361 GTCCATTGAGGATGACCATCTTAATTAAGTATGTAAGTATGTAAGTATGTAAGTATGTA 420  
421 TGGGAAAAGGGGATGGGAGTCTCTAGATTAAAGTCTGAGGCGAGTCTATTTCTCAA 480  
421 TGGGAAAAGGGGATGGGAGTCTCTAGATTAAAGTCTGAGGCGAGTCTATTTCTCAA 480  
421 TGGGAAAAGGGGATGGGAGTCTCTAGATTAAAGTCTGAGGCGAGTCTATTTCTCAA 480  
481 TTTGATTTCATATGAAAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGGACTGTGACT 540  
481 TTTGATTTCATATGAAAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGGACTGTGACT 540  
481 TTTGATTTCATATGAAAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGGACTGTGACT 540  
481 TTTGATTTCATATGAAAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGGACTGTGACT 540  
541 GAAGACGTGACGGCTTATTAACACTGGCACTTATTAACACTTATTAACACTGGCAACAGG 600  
541 GAAGACGTGACGGCTTATTAACACTGGCACTTATTAACACTTATTAACACTGGCAACAGG 600  
601 CGTTGAGGTTGAAGATCACTTTCAAAACACAGAAAGAGTGTGCTGCTCTGACG 660  
601 CGTTGAGGTTGAAGATCACTTTCAAAACACAGAAAGAGTGTGCTGCTCTGACG 660  
601 CGTTGAGGTTGAAGATCACTTTCAAAACACAGAAAGAGTGTGCTGCTCTGACG 660  
661 GTACGAGCACTGGCTGCAAGAGAGTATTTAGTGAAGCTACCTTCAACAAATCTTT 720  
661 GTACGAGCACTGGCTGCAAGAGAGTATTTAGTGAAGCTACCTTCAACAAATCTTT 720  
661 GTACGAGCACTGGCTGCAAGAGAGTATTTAGTGAAGCTACCTTCAACAAATCTTT 720  
721 GCACTTTACATACAGCTGTCAATATGCTAACTCCCTAGTCCACAGATGGCTGTACACA 780  
721 GCACTTTACATACAGCTGTCAATATGCTAACTCCCTAGTCCACAGATGGCTGTACACA 780  
781 CTCGTTTCTGCTTCCCATCTGTTGACATTTGTCAAGAACCGAATAATTAGAAATGTGGGT 840  
781 CTCGTTTCTGCTTCCCATCTGTTGACATTTGTCAAGAACCGAATAATTAGAAATGTGGGT 840

841 ATTTATTTGCTGCTGAGGACCAATCCAGGCTTTTTCATATTTACAGGACATGCTTAC 900  
841 ATTTATTTGCTGCTGAGGACCAATCCAGGCTTTTTCATATTTACAGGACATGCTTAC 900  
901 TTAATGGGCTATCTTCCACAGGTTTGAACCATTTGTTTATTTATTTATTTATTTGCT 960  
901 TTAATGGGCTATCTTCCACAGGTTTGAACCATTTGTTTATTTATTTATTTATTTGCT 960  
961 GCATGAGGTAGGATGATATACGATATGATAGGATGATGATGATGATGATGATGATGAT 1020  
961 GCATGAGGTAGGATGATATACGATATGATAGGATGATGATGATGATGATGATGATGAT 1020  
1021 ATCATTTGACATATCCCAAGCAAGTACAGGCTTTGTAAGTGTATGTTGAGGAGT 1080  
1021 ATCATTTGACATATCCCAAGCAAGTACAGGCTTTGTAAGTGTATGTTGAGGAGT 1080  
1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGACCCAGTGGCCCTTAACATGGGACCACT 1140  
1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGACCCAGTGGCCCTTAACATGGGACCACT 1140  
1141 CTAGGCTTAAGGTATCTTTAGTTTAAATATATATATATATATATATATATATATAT 1200  
1141 CTAGGCTTAAGGTATCTTTAGTTTAAATATATATATATATATATATATATATATAT 1200  
1201 ACAGGCTTTAATCCACAGACCTTACAGAGGCTGAGAGGCTGAGATATATACACAGG 1260  
1201 ACAGGCTTTAATCCACAGACCTTACAGAGGCTGAGAGGCTGAGATATATACACAGG 1260  
1261 CTGGGGTCAAGGCTGGGCTGTTTCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1320  
1261 CTGGGGTCAAGGCTGGGCTGTTTCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1320  
1321 CTTGCTGTATATGCTGTCAGAGGCTGTCAGATCCCTTGGAGCTGAGCTTAAAGACG 1380  
1321 CTTGCTGTATATGCTGTCAGAGGCTGTCAGATCCCTTGGAGCTGAGCTTAAAGACG 1380  
1321 CTTGCTGTATATGCTGTCAGAGGCTGTCAGATCCCTTGGAGCTGAGCTTAAAGACG 1380  
1381 TGATCAGGCTGCTTACAGATGCTGGAATATGAACCCAGGCTGCTTACAGAGGAGC 1440  
1381 TGATCAGGCTGCTTACAGATGCTGGAATATGAACCCAGGCTGCTTACAGAGGAGC 1440  
1381 TGATCAGGCTGCTTACAGATGCTGGAATATGAACCCAGGCTGCTTACAGAGGAGC 1440  
1441 AGTCTCTTAACTTCTGAGCCACCCCTTCAACCCCTGTTTAAAGACTCTTAACTTAACT 1500  
1441 AGTCTCTTAACTTCTGAGCCACCCCTTCAACCCCTGTTTAAAGACTCTTAACTTAACT 1500  
1441 AGTCTCTTAACTTCTGAGCCACCCCTTCAACCCCTGTTTAAAGACTCTTAACTTAACT 1500  
1501 TGTATGTGGGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1501 TGTATGTGGGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1561 CTGAGCCCTTACCCACAGACCTAGTATGATGATGATGATGATGATGATGATGATGAT 1620  
1561 CTGAGCCCTTACCCACAGACCTAGTATGATGATGATGATGATGATGATGATGATGAT 1620  
1621 ATAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
1621 ATAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
1681 TTTAGCTGCTGCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
1681 TTTAGCTGCTGCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
1741 GAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
1741 GAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
1801 AATAGTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
1801 AATAGTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
1861 GAGATGCTCAGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920  
1861 GAGATGCTCAGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920  
1921 AGGACTATATATGCTGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980

Db 1921 AGAGACATATGTGGCTCACAGCCATCTGTAATCCAGTCCAGAGGGTCCACACCTT 1980  
QY 1981 CTCTGGCCCTCCACAGGACACATCATATGTATACAGACATCATGACGAGAAACACC 2040  
Db 1981 CTCTGGCCCTCCACAGGACACATCATATGTATACAGACATCATGACGAGAAACACC 2040  
QY 2041 CATACACATTAATTAATTAAGAACTTAAAGGTGATGTGTGTAAACATTTGCT 2100  
Db 2041 CATACACATTAATTAATTAAGAACTTAAAGGTGATGTGTGTAAACATTTGCT 2100  
QY 2101 TACACATGCTGATTTGAAGACATATACAGCCACACACTGAAGGGATCTGGGGCTGAG 2160  
Db 2101 TACACATGCTGATTTGAAGACATATACAGCCACACACTGAAGGGATCTGGGGCTGAG 2160  
QY 2161 AGATGGCTGAGGGTTAAGGACATGCTCTCCGAAAGGAAAGGCTCCGATTCAAAT 2220  
Db 2161 AGATGGCTGAGGGTTAAGGACATGCTCTCCGAAAGGAAAGGCTCCGATTCAAAT 2220  
QY 2221 CCTAGCACCATGGTGGCTCACAAACCATCCATAATGAGATCTGACACCCTTTGCT 2280  
Db 2221 CCTAGCACCATGGTGGCTCACAAACCATCCATAATGAGATCTGACACCCTTTGCT 2280  
QY 2281 GCATCTGAAGACAGCTGACAGAGCTACAGTGTACTTGAATTAATTAATCTTTT 2340  
Db 2281 GCATCTGAAGACAGCTGACAGAGCTACAGTGTACTTGAATTAATTAATCTTTT 2340  
QY 2341 TTTAAAAAATGAAGGGGATCTGAGACACCTCAAAAGAGATTAATGACGATACACAG 2400  
Db 2341 TTTAAAAAATGAAGGGGATCTGAGACACCTCAAAAGAGATTAATGACGATACACAG 2400  
QY 2401 GGTGATTAATCTATCTGAGATTTTTCCTTCCGTTGGCTTGCACACTGGGTGAGACAG 2460  
Db 2401 GGTGATTAATCTATCTGAGATTTTTCCTTCCGTTGGCTTGCACACTGGGTGAGACAG 2460  
QY 2461 CCCCTTTTCAATTACAGAAAGGGGTCTACATTTTCTGACAAAACAGACCTGCACT 2520  
Db 2461 CCCCTTTTCAATTACAGAAAGGGGTCTACATTTTCTGACAAAACAGACCTGCACT 2520  
QY 2521 ATGTTACTGTCTGCTGACTATGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2580  
Db 2521 ATGTTACTGTCTGCTGACTATGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2580  
QY 2581 ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGTCA 2640  
Db 2581 ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGTCA 2640  
QY 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTATCTTCCCTCTTAAAGAGACACGATTC 2700  
Db 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTATCTTCCCTCTTAAAGAGACACGATTC 2700  
QY 2701 AAGGTGGCAGAAATCTACAGGGGGCAGAGGCGGAGGGGAAAGCAGGCGATGTTTC 2760  
Db 2701 AAGGTGGCAGAAATCTACAGGGGGCAGAGGCGGAGGGGAAAGCAGGCGATGTTTC 2760  
QY 2761 AGAGACCTACAGCAGAGGGCAGCAGAGAGTCCAGGTCCAGGGCAGGGAGGTGAG 2820  
Db 2761 AGAGACCTACAGCAGAGGGCAGCAGAGAGTCCAGGTCCAGGGCAGGGAGGTGAG 2820  
QY 2821 CCTTTTTCGAGGAGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2880  
Db 2821 CCTTTTTCGAGGAGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2880  
QY 2881 GGTCAATAAAAGTGGAGGTGTGCTCAGTCAAGAAAGAGGAAAGGAAAGGCGCTTGT 2940  
Db 2881 GGTCAATAAAAGTGGAGGTGTGCTCAGTCAAGAAAGAGGAAAGGAAAGGCGCTTGT 2940  
QY 2941 GCCCACTGAGCGAGGGTCACTGAGTGAAGAGATCTCAGAGGGGTGCCAGAGGCCAC 3000  
Db 2941 GCCCACTGAGCGAGGGTCACTGAGTGAAGAGATCTCAGAGGGGTGCCAGAGGCCAC 3000  
QY 3001 CTGTCTGTCCCAAGGAAACCCCAAGTGTAACTCTGGCTGTGAGTTCCAGCTA 3060  
Db 3001 CTGTCTGTCCCAAGGAAACCCCAAGTGTAACTCTGGCTGTGAGTTCCAGCTA 3060

Db 3001 CTGTCTGTCCCAAGGAAACCCCAAGTGTAACTCTGGCTGTGAGTTCCAGCTA 3060  
QY 3061 CAAGACCCCGAGAGTCTTACTCATCCATCCAGTGGCCCCCGGCGGCGGCGGCGGCGG 3120  
Db 3061 CAAGACCCCGAGAGTCTTACTCATCCATCCAGTGGCCCCCGGCGGCGGCGGCGGCGG 3120  
QY 3121 CCCCCGACTCCCTGGCAGCTTCTAGGGCTGGAGGGTGGCCAGCCCTGGTGGGGTTC 3180  
Db 3121 CCCCCGACTCCCTGGCAGCTTCTAGGGCTGGAGGGTGGCCAGCCCTGGTGGGGTTC 3180  
QY 3181 CTACCTGACAGTGTAGAGCCCGAGGTCTAGCCGGAAGTGCACCCCATCTGAACTGCAGA 3240  
Db 3181 CTACCTGACAGTGTAGAGCCCGAGGTCTAGCCGGAAGTGCACCCCATCTGAACTGCAGA 3240  
QY 3241 GCCAAGGGGGGCGACAGGGCAGCTCAGGCTGACAGGCTTGTGCTGGGCTGAGTTCC 3300  
Db 3241 GCCAAGGGGGGCGACAGGGCAGCTCAGGCTGACAGGCTTGTGCTGGGCTGAGTTCC 3300  
QY 3301 AGGACCTGGGCGACACTTCCACCCACCCCATCCATCTCTGAGGGCCCTATCTTC 3360  
Db 3301 AGGACCTGGGCGACACTTCCACCCACCCCATCCATCTCTGAGGGCCCTATCTTC 3360  
QY 3361 CTATATGCTGAAGAAAGTCTGAGGGGGGGGGGTGTGTGAGGACAAAGTCTTTC 3420  
Db 3361 CTATATGCTGAAGAAAGTCTGAGGGGGGGGGGTGTGTGAGGACAAAGTCTTTC 3420  
QY 3421 GTCTCTGACAGGCTGGCACAAGTCTCTAGATCTCCAGGTGGTGGCTCTTC 3480  
Db 3421 GTCTCTGACAGGCTGGCACAAGTCTCTAGATCTCCAGGTGGTGGCTCTTC 3480  
QY 3481 CAGACAGGTAAAGCAATTTGGGTGGGACACATGGTGCACACAGGTGTGGAGGGACAG 3540  
Db 3481 CAGACAGGTAAAGCAATTTGGGTGGGACACATGGTGCACACAGGTGTGGAGGGACAG 3540  
QY 3541 GGTCTTGTCTCTCTGCGACGCTGCTTCTGTAGACACTTGGTATAGTTTGGGG 3600  
Db 3541 GGTCTTGTCTCTCTGCGACGCTGCTTCTGTAGACACTTGGTATAGTTTGGGG 3600  
QY 3601 TGAAGTAAAGTGTCTGAAACCTGAAAGAAAGCAAGAGCAGCAGGCTTCTGGGCT 3660  
Db 3601 TGAAGTAAAGTGTCTGAAACCTGAAAGAAAGCAAGAGCAGCAGGCTTCTGGGCT 3660  
QY 3661 TCAATGAAGAAATTCACAGACCCCTTCTCTGTAAGTACCTTCATCTGATG 3720  
Db 3661 TCAATGAAGAAATTCACAGACCCCTTCTCTGTAAGTACCTTCATCTGATG 3720  
QY 3721 ATTCCCTGGGACCAAGTGGCTCTGAGACTGAGATTTCTACAAATTAATCGAGACAG 3780  
Db 3721 ATTCCCTGGGACCAAGTGGCTCTGAGACTGAGATTTCTACAAATTAATCGAGACAG 3780  
QY 3781 CCTGAGACTTGGACCTCGTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3840  
Db 3781 CCTGAGACTTGGACCTCGTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3840  
QY 3841 TCAATGCTTAACACATCTGAATATGTTTGTGTGACCAATTCCTGACACTCTGGGA 3900  
Db 3841 TCAATGCTTAACACATCTGAATATGTTTGTGTGACCAATTCCTGACACTCTGGGA 3900  
QY 3901 GGTGATATCTTGGCAGATGTATCTGGGATGTAACTGAGCCACAGAGAGAGGGG 3960  
Db 3901 GGTGATATCTTGGCAGATGTATCTGGGATGTAACTGAGCCACAGAGAGAGAGGGG 3960  
QY 3961 AGAGTGAAGAGCTGTCTCTAGAGCCCTATTAAGGCTGAGACATCCCTTCTTCAAAAT 4020  
Db 3961 AGAGTGAAGAGCTGTCTCTAGAGCCCTATTAAGGCTGAGACATCCCTTCTTCAAAAT 4020  
QY 4021 GGCCCTCATATTTTGGTTACCATATCTATTTATACAGAGTGGGAGGAAAGCA 4080  
Db 4021 GGCCCTCATATTTTGGTTACCATATCTATTTATACAGAGTGGGAGGAAAGCA 4080  
QY 4081 AACCTGCCAAGAAAGTTGGAGCTCACTCAGACCAAGGTTATGTCTCAAGAAATCCCCG 4140  
Db 4081 AACCTGCCAAGAAAGTTGGAGCTCACTCAGACCAAGGTTATGTCTCAAGAAATCCCCG 4140

OY 4141 TCACCTTAGGTTGGGAGAAATCTGCTCTGGGGGCTTCCAGGTCCTTGCTTAGCAGAGGCT 4200  
|||||  
Db 4141 TCACCTTAGGTTGGGAGAAATCTGCTCTGGGGGCTTCCAGGTCCTTGCTTAGCAGAGGCT 4200  
OY 4201 ATCCTTGTATAGGGCAATGACCTAGCTATGCTGTACTACATCCCTGCTCCAGTTAAAG 4260  
|||||  
Db 4201 ATCCTTGTATAGGGCAATGACCTAGCTATGCTGTACTACATCCCTGCTCCAGTTAAAG 4260  
OY 4261 CTGGAAGTAAACCAGGCGAGCCCGAGATTCCTACAGTTGTACCCCAAGAACAA 4320  
|||||  
Db 4261 CTGGAAGTAAACCAGGCGAGCCCGAGATTCCTACAGTTGTACCCCAAGAACAA 4320  
OY 4321 AGACAGTAGATATCCAAAGATAGTAGCTGGGAGAAAGAACTTAAACCCCAAG 4380  
|||||  
Db 4321 AGACAGTAGATATCCAAAGATAGTAGCTGGGAGAAAGAACTTAAACCCCAAG 4380  
OY 4381 GCCCAGAGGTTCCGCTCCCTAGTCCACAAATGCCAGTATGATGCTAGCTACTATGGGCTG 4440  
|||||  
Db 4381 GCCCAGAGGTTCCGCTCCCTAGTCCACAAATGCCAGTATGATGCTAGCTACTATGGGCTG 4440  
OY 4441 TGAGTTGGTAGCTACAAAGCATGATGATGTCATGTCGTGTGTATATATCTGAGAC 4500  
|||||  
Db 4441 TGAGTTGGTAGCTACAAAGCATGATGATGTCATGTCGTGTGTATATATCTGAGAC 4500  
OY 4501 TTGGGAGGCTGAGCAGAGAGATTCATATGTTTGAGGCCAGCCTGAGCTATAGAGCA 4560  
|||||  
Db 4501 TTGGGAGGCTGAGCAGAGAGATTCATATGTTTGAGGCCAGCCTGAGCTATAGAGCA 4560  
OY 4561 GACCTTTCCTTTAGAAAAAATGAAAGCCAGAGGTCGAGAGGCTGTCAGGCTTTATCCCA 4620  
|||||  
Db 4561 GACCTTTCCTTTAGAAAAAATGAAAGCCAGAGGTCGAGAGGCTGTCAGGCTTTATCCCA 4620  
OY 4621 GCACCTTGGGAGGCAAGAGCAGATTTCTGAGTTCAGAGCCAGCCTGCTATAGAGT 4680  
|||||  
Db 4621 GCACCTTGGGAGGCAAGAGCAGATTTCTGAGTTCAGAGCCAGCCTGCTATAGAGT 4680  
OY 4681 GAGTTTCCAGCAGACCCAGGCTTACACAGAAACCCCTGTTTGAAGAAACCGAAGAAACAA 4740  
|||||  
Db 4681 GAGTTTCCAGCAGACCCAGGCTTACACAGAAACCCCTGTTTGAAGAAACCGAAGAAACAA 4740  
OY 4741 AACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 4800  
|||||  
Db 4741 AACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 4800  
OY 4801 TAGGCTGTCTCTCTAGTGTGTAGAGTTTGGGAGCTTCAGACTTATATATATATATAGGCC 4860  
|||||  
Db 4801 TAGGCTGTCTCTCTAGTGTGTAGAGTTTGGGAGCTTCAGACTTATATATATATATAGGCC 4860  
OY 4861 TTTTATATCTGCTGACAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTTGAGA 4920  
|||||  
Db 4861 TTTTATATCTGCTGACAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTTGAGA 4920  
OY 4921 AAGTACCTCTGGCAGGCAAAATTCGGAAGGCTTCGGAAGAGTGTGCCGAT 4980  
|||||  
Db 4921 AAGTACCTCTGGCAGGCAAAATTCGGAAGGCTTCGGAAGAGTGTGCCGAT 4980  
OY 4981 CAGACTACTGTTCTAGAAAGCAGAAAGAGGTTGGAAGATGTTGTGTGACAGACAGTT 5040  
|||||  
Db 4981 CAGACTACTGTTCTAGAAAGCAGAAAGAGGTTGGAAGATGTTGTGTGACAGACAGTT 5040  
OY 5041 GGAACAGAAAGACAGAGGGGAGGCAATCCAAAGATTCGAAACATGTACTACTTTGGT 5100  
|||||  
Db 5041 GGAACAGAAAGACAGAGGGGAGGCAATCCAAAGATTCGAAACATGTACTACTTTGGT 5100  
OY 5101 TCTTGGGTGACAAGTGTCCCGAGGATAGGCTGTAGAAAGGGGACAGAGGGTGAGCC 5160  
|||||  
Db 5101 TCTTGGGTGACAAGTGTCCCGAGGATAGGCTGTAGAAAGGGGACAGAGGGTGAGCC 5160  
OY 5161 AATGAGTTCAAGTTGAGGAGACATCCAGCCAGGCTCTTGTGCGGACAGTTAAAGATG 5220  
|||||  
Db 5161 AATGAGTTCAAGTTGAGGAGACATCCAGCCAGGCTCTTGTGCGGACAGTTAAAGATG 5220

OY 5221 AGAGCCCTCTAACCCCTCCCTGAAGTTTAGGGGAGACAGAGAGCTGAGAGATCTCTTA 5280  
|||||  
Db 5221 AGAGCCCTCTAACCCCTCCCTGAAGTTTAGGGGAGACAGAGAGCTGAGAGATCTCTTA 5280  
OY 5281 GGGTGAAGAGAGATATCTGCTGACCAACATGGCTAGAGCAGAAAGATTTGGACAG 5340  
|||||  
Db 5281 GGGTGAAGAGAGATATCTGCTGACCAACATGGCTAGAGCAGAAAGATTTGGACAG 5340  
OY 5341 TTACCCCTCAGAACACACATCCCTCTTGAGCTTAAGAGAGGCTGGGCCCTTTCTGTTT 5400  
|||||  
Db 5341 TTACCCCTCAGAACACACATCCCTCTTGAGCTTAAGAGAGGCTGGGCCCTTTCTGTTT 5400  
OY 5401 AAGAATCTTACTTTCTTCAGAGAGAGGAGCAGACCTTTGTCCCTCCCTGTGGTCA 5460  
|||||  
Db 5401 AAGAATCTTACTTTCTTCAGAGAGAGGAGCAGACCTTTGTCCCTCCCTGTGGTCA 5460  
OY 5461 TAAACACCCCTGTGTGTAACTTGTATTTTACTGTACATGTTGTCTCCAGAGACGTCA 5520  
|||||  
Db 5461 TAAACACCCCTGTGTGTAACTTGTATTTTACTGTACATGTTGTCTCCAGAGACGTCA 5520  
OY 5521 TCTGTAGACCTCTGCTCTTAACCTCACCAAGATAGGCCACATTTCTCACCAAGAG 5580  
|||||  
Db 5521 TCTGTAGACCTCTGCTCTTAACCTCACCAAGATAGGCCACATTTCTCACCAAGAG 5580  
OY 5581 TGCAAGAGAGGCTTAGAGAAAGGTTAACAGTAAACAAAGTGGCCAGAAATAAACAAA 5640  
|||||  
Db 5581 TGCAAGAGAGGCTTAGAGAAAGGTTAACAGTAAACAAAGTGGCCAGAAATAAACAAA 5640  
OY 5641 ACTACTATCCCTTTTACCCAAATGGTTTGTCTGAACAGAGAGGCTGTGTAGTAT 5700  
|||||  
Db 5641 ACTACTATCCCTTTTACCCAAATGGTTTGTCTGAACAGAGAGGCTGTGTAGTAT 5700  
OY 5701 GT 5760  
|||||  
Db 5701 GT 5760  
OY 5761 CTTGGGGAGCTTTTCAATGCTTAAGAAATTCGATTTGGGGCCATGCAACAGGGAT 5820  
|||||  
Db 5761 CTTGGGGAGCTTTTCAATGCTTAAGAAATTCGATTTGGGGCCATGCAACAGGGAT 5820  
OY 5821 TGGGAGAGTACAGCTTCTGCAAAACACAGTAAAGTGGCCAGAGATGATTTGGTGGCTGAA 5880  
|||||  
Db 5821 TGGGAGAGTACAGCTTCTGCAAAACACAGTAAAGTGGCCAGAGATGATTTGGTGGCTGAA 5880  
OY 5881 TCACCAAGGGGAGGCTGATCAGATGAGAGACAGAAATCACAAGATTAAGCCACCTGTGGG 5940  
|||||  
Db 5881 TCACCAAGGGGAGGCTGATCAGATGAGAGACAGAAATCACAAGATTAAGCCACCTGTGGG 5940  
OY 5941 GCTCAGAAAGGGAGTTTACAAGAGGTTAAAGGCCAAGCCATTTTATCCAAAGACATGAC 6000  
|||||  
Db 5941 GCTCAGAAAGGGAGTTTACAAGAGGTTAAAGGCCAAGCCATTTTATCCAAAGACATGAC 6000  
OY 6001 TCAAAATCAAAAGTCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGACACC 6060  
|||||  
Db 6001 TCAAAATCAAAAGTCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGACACC 6060  
OY 6061 TGACCTTGACTTATATGTCATAGGCCAAGGACAGTCAACAGAGGTTGACTGGCTCTTA 6120  
|||||  
Db 6061 TGACCTTGACTTATATGTCATAGGCCAAGGACAGTCAACAGAGGTTGACTGGCTCTTA 6120  
OY 6121 CTGAGCTTGGAGGAGGACGTTGAGATGGGTGACCTCCATCCGATGAGAGAGGCTGAG 6180  
|||||  
Db 6121 CTGAGCTTGGAGGAGGACGTTGAGATGGGTGACCTCCATCCGATGAGAGAGGCTGAG 6180  
OY 6181 CACCACAGGTACAAAGTGTCCCTGTCTCATGCCAGAGATTCCTGGCCAGTTTTCAGAG 6240  
|||||  
Db 6181 CACCACAGGTACAAAGTGTCCCTGTCTCATGCCAGAGATTCCTGGCCAGTTTTCAGAG 6240  
OY 6241 GACTTAAGACTCATCTGTGTGGAACAAGTATCCAAAGCCCTTAAGCCCTTTTGGTCT 6300  
|||||  
Db 6241 GACTTAAGACTCATCTGTGTGGAACAAGTATCCAAAGCCCTTAAGCCCTTTTGGTCT 6300  
OY 6301 AATTAAATCAGAACCCCTGGGAGTGACAGGCTGTAGACAGAGAGCTTTTAAAGACCTC 6360  
|||||



```
|||||
Db 6301 AATTAAATCAGAACCCCTGGGGATGAGGCTCTAGAGCAGCAGGAGCTTTTAAAACTC 6360
Qy 6361 CCAGGTATTCCTGATCAGCAGCTGGAAACAAACAGCTACAGGTTCAAACGAAGAGGC 6420
Db 6361 CCAGGTATTCCTGATCAGCAGCTGGAAACAAACAGCTACAGGTTCAAACGAAGAGGC 6420
Qy 6421 AAAGCTAGGGAACCTTGGATGGGAGCCCTTCTCCAGGCCAGTAGATGGAGGCTGTT 6480
Db 6421 AAAGCTAGGGAACCTTGGATGGGAGCCCTTCTCCAGGCCAGTAGATGGAGGCTGTT 6480
Qy 6481 AGCAGTGGTGGCAGCTTCTCTGCTGTCATATAGTATCATCATCAGCTATCCATCAT 6540
Db 6481 AGCAGTGGTGGCAGCTTCTCTGCTGTCATATAGTATCATCATCAGCTATCCATCAT 6540
Qy 6541 ACACCCAGCCATCATTTATGACCCACCCATCCATCATCATCATCATCATCATCATCAT 6600
Db 6541 ACACCCAGCCATCATTTATGACCCACCCATCCATCATCATCATCATCATCATCATCAT 6600
Qy 6601 CCAGCATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Db 6601 CCAGCATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Qy 6661 CATTTATCCAAAGAGAACTGGTATGTACTAAATGGGAGATTTAATTAATTTTGA 6720
Db 6661 CATTTATCCAAAGAGAACTGGTATGTACTAAATGGGAGATTTAATTAATTTTGA 6720
Qy 6721 AGGCTGTGATGACGATTTGACATGTATGACATGTGACAGTATACACAGCAGCAGCTG 6780
Db 6721 AGCTCTGTGATGACGATTTGACATGTATGACATGTGACAGTATACACAGCAGCAGCTG 6780
Qy 6781 TGGCAATCGGAGAAAGTTTGGGTGTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
Db 6781 TGGCAATCGGAGAAAGTTTGGGTGTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
Qy 6841 TTGACATCAATTTATCGGGCTGTGCAAGTGTCTTTACACGAGCCATTTCTCTGACA 6900
Db 6841 TTGACATCAATTTATCGGGCTGTGCAAGTGTCTTTACACGAGCCATTTCTCTGACA 6900
Qy 6901 CATCATTTATTTAGAAAGCATCTTATGTATGTATCCAGGCTGAGCTTGCATATGCG 6960
Db 6901 CATCATTTATTTAGAAAGCATCTTATGTATGTATCCAGGCTGAGCTTGCATATGCG 6960
Qy 6961 CCAGGATGACCTTTAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020
Db 6961 CCAGGATGACCTTTAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020
Qy 7021 GTTCACTGGTGAATGCTTTTATCCAGCAGCTGTGAGGGGGGGGGGGGGGAGGAGTC 7080
Db 7021 GTTCACTGGTGAATGCTTTTATCCAGCAGCTGTGAGGGGGGGGGGGGGGAGGAGTC 7080
Qy 7081 CTTGAGTGGAGGCTGTTGGTGTACAGAGTTCAAGATACCTGGGGCTTTTACAGGAA 7140
Db 7081 CTTGAGTGGAGGCTGTTGGTGTACAGAGTTTCAAGATACCTGGGGCTTTTACAGGAA 7140
Qy 7141 ACCCTATCCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200
Db 7141 ACCCTATCCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200
Qy 7201 TTAGAGATTTAGTAGGAGTAGTAGGAGCTGGTGAGGAGAGTCACTCTTTCTTTGTAAT 7260
Db 7201 TTAGAGATTTAGTAGGAGTAGTAGGAGCTGGTGAGGAGAGTCACTCTTTCTTTGTAAT 7260
Qy 7261 AATATAGTAAAGTCTACAAAGATGATATATATATATATATATATATATATATATATATAT 7320
Db 7261 AATATAGTAAAGTCTACAAAGATGATATATATATATATATATATATATATATATATATAT 7320
Qy 7321 TATCTATCTATCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAG 7380
Db 7321 TATCTATCTATCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAG 7380
Qy 7381 CTGCTTGAAGTCTGAATGCTCTATTTCTGGGTCAGCTCTACCCCTAGTGTGGGTTT 7440
Db 7381 CTGCTTGAAGTCTGAATGCTCTATTTCTGGGTCAGCTCTACCCCTAGTGTGGGTTT 7440

|||||
Db 7381 CTGCTTGAAGTCTGAATGCTCTATTTCTGGGTCAGCTCTACCCCTAGTGTGGGTTT 7440
Qy 7441 ACCAAGCCCGACATTTATTTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTT 7500
Db 7441 ACCAAGCCCGACATTTATTTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTATTT 7500
Qy 7501 GGACTCAGGGCTCTGTGCATGCTTAAGCAAGCTCTCTGCAACAGAGCTGACAGCTCAGTCC 7560
Db 7501 GGACTCAGGGCTCTGTGCATGCTTAAGCAAGCTCTCTGCAACAGAGCTGACAGCTCAGTCC 7560
Qy 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTCTATTTTGGAGGCTATGAGTCTCTC 7620
Db 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTCTATTTTGGAGGCTATGAGTCTCTC 7620
Qy 7621 CACCTCCAGTTCAGAGCTTTGTCATATCCAGTGGGGGGGAGCCTGTGTCACCA 7680
Db 7621 CACCTCCAGTTCAGAGCTTTGTCATATCCAGTGGGGGGGAGCCTGTGTCACCA 7680
Qy 7681 GTGCCCTGTTCCCTGTCTCAGACCTACATATTTGGCCGTGTGAAAGTTCATGTAATG 7740
Db 7681 GTGCCCTGTTCCCTGTCTCAGACCTACATATTTGGCCGTGTGAAAGTTCATGTAATG 7740
Qy 7741 GGATGCGTCTGTGATATCTTTTATGAGCTGGCCCTTTATCTTATGACAGCTTGTGTG 7800
Db 7741 GGATGCGTCTGTGATATCTTTTATGAGCTGGCCCTTTATCTTATGACAGCTTGTGTG 7800
Qy 7801 GGCCATGTCACGATGATACATCTATGATATCATCATCTTATGAGCTTATGATGCTTCT 7860
Db 7801 GGCCATGTCACGATGATACATCTATGATATCATCATCTTATGAGCTTATGATGCTTCT 7860
Qy 7861 TGTGTGATTAACACACTTCTGTTTCAATTTACTGATGAAATTTGTGGCCACACCCAC 7920
Db 7861 TGTGTGATTAACACACTTCTGTTTCAATTTACTGATGAAATTTGTGGCCACACCCAC 7920
Qy 7921 CTTTTTTTTTTTTTTTATTTGAGCAAGCTTCTGTGATATTTGCAATCTTGGCTGTCT 7980
Db 7921 CTTTTTTTTTTTTTTTATTTGAGCAAGCTTCTGTGATATTTGCAATCTTGGCTGTCT 7980
Qy 7981 GAGCTCACTGTGTAGACAGGCTGTGAGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8040
Db 7981 GAGCTCACTGTGTAGACAGGCTGTGAGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8040
Qy 8041 AGAGTACCATGAACTTCAAGCAATTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 8100
Db 8041 AGAGTACCATGAACTTCAAGCAATTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 8100
Qy 8101 TGTATGCGTGTATATGTGCATGTTTGTCTTCAAGGCTGTACATGTGTGACCTGTG 8160
Db 8101 TGTATGCGTGTATATGTGCATGTTTGTCTTCAAGGCTGTACATGTGTGACCTGTG 8160
Qy 8161 TGGGACAGAGAAACCAACCGATGTGCATCTCTCAGATACATCGCATCTGTTAATATGTA 8220
Db 8161 TGGGACAGAGAAACCAACCGATGTGCATCTCTCAGATACATCGCATCTGTTAATATGTA 8220
Qy 8221 TGTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 8280
Db 8221 TGTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 8280
Qy 8281 CTTCCCTTGTGGCTCTCCGATTAACCTACGCTCTCGGGCTAGTAGAGCATGCTTCA 8340
Db 8281 CTTCCCTTGTGGCTCTCCGATTAACCTACGCTCTCGGGCTAGTAGAGCATGCTTCA 8340
Qy 8341 CTGATGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400
Db 8341 CTGATGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400
Qy 8401 CGCAGCTCAGTGGCTAAAGCTACCAAGTATCAGAGTGGCTAGCCAGGAGACTCAG 8460
Db 8401 CGCAGCTCAGTGGCTAAAGCTACCAAGTATCAGAGTGGCTAGCCAGGAGACTCAG 8460
Qy 8461 GGATATGCTGGCTCTGCTCTCAGAGTGTAGAAATTCAGGCAATACATCTGCTGGAAG 8520
Db 8461 GGATATGCTGGCTCTGCTCTCAGAGTGTAGAAATTCAGGCAATACATCTGCTGGAAG 8520
```

```

Oy 8521 ATTTTAACCTGAATCCTGAGATAGACAGGACCTACCAATGAGGCTTTTGT 8580
Db 8521 ATTTTAACCTGAATCCTGAGATAGACAGGACCTACCAATGAGGCTTTTGT 8580
Oy 8581 GTTTGGTTGGTTCTCTGCATAGATACAGGACCTGCAATATAGTACCTGGCTAC 8640
Db 8581 GTTTGGTTGGTTCTCTGCATAGATAGACAGGACCTGCAATATAGTACCTGGCTAC 8640
Oy 8641 ATACATCTTGTCTCAAAAAGCCTATAGAGTAGGAGGAGCTGAAGAAGACCTT 8700
Db 8641 ATACATCTTGTCTCAAAAAGCCTATAGAGTAGGAGGAGCTGAAGAAGACCTT 8700
Oy 8701 AAGCGCGCTGTATAGACACAGATAGCCTGCACATATATAGCAAGACCTTTTCAAA 8760
Db 8701 AAGCGCGCTGTATAGACACAGATAGCCTGCACATATATAGCAAGACCTTTTCAAA 8760
Oy 8761 ACATGAGAGGAGGGATATGTTTAAGTCTGGGCTGTGTACAGCAGCTAAGGAGCCAA 8820
Db 8761 ACATGAGAGGAGGGATATGTTTAAGTCTGGGCTGTGTACAGCAGCTAAGGAGCCAA 8820
Oy 8821 TGTAGACATTTGACTATAGAAAGATCATCATCAAAAGCCGGGTGGGAGGTTGG 8880
Db 8821 TGTAGACATTTGACTATAGAAAGATCATCATCAAAAGCCGGGTGGGAGGTTGG 8880
Oy 8881 ACTACAGTGTCAAGACCCCATAGAGAGCCAGTTTCCCTTCTCTGGGCTCAAGC 8940
Db 8881 ACTACAGTGTCAAGACCCCATAGAGAGCCAGTTTCCCTTCTCTGGGCTCAAGC 8940
Oy 8941 CTGGCTCAGCGCCACTCTCTACATGCCCTTCTCTAGGCTGTCCACCATG 8995
Db 8941 CTGGCTCAGCGCCACTCTCTACATGCCCTTCTCTAGGCTGTCCACCATG 8995

```

```

RESULT 2
LOCUS AC098570 192060 bp DNA linear HTG 20-AUG-2002
DEFINITION Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered
ACCESSION AC098570.2 GI:22325297
VERSION AC098570.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192060)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-278N11
Unpublished
2 (bases 1 to 192060)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nauyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, R., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (20-OCT-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 192060)  
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Chepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Fato, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
Gaidyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,  
Murphy, T., Naylor, J., Nauyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 20, 2002 this sequence version replaced gi:16356898.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: 278\_N\_11  
Center clone name: 144796  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960721  
Consensus quality: 190464 bases at least Q40  
Consensus quality: 191061 bases at least Q40  
Consensus quality: 191345 bases at least Q20  
Insert size: 163000; agarose-1p  
Insert size: 191560; sum-of-contigs  
Quality coverage: 12.4 in Q20 bases; agarose-1p  
Quality coverage: 10.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 4921: contig of 4921 bp in length  
\* 4922 5021: gap of 100 bp  
\* 5022 15200: contig of 10179 bp in length  
\* 15201 15300: gap of 100 bp  
\* 15301 78692: contig of 63392 bp in length  
\* 78693 78792: gap of 100 bp  
\* 78793 116118: contig of 37326 bp in length  
\* 116119 116218: gap of 100 bp  
\* 116219 174491: contig of 58273 bp in length  
\* 174492 174591: gap of 100 bp  
\* 174592 192060: contig of 17469 bp in length.  
Location/Qualifiers  
1. 192060  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"

```

/clone="RP23-278N11"
/clone_lib="RPCT-23 Female Mouse BAC"
1..4921
/note="assembly_fragment
clone_end:SP6
vector_side:left"
5022..15200
/note="assembly_fragment"
15301..78692
/note="assembly_fragment"
78793..116118
/note="assembly_fragment"
116219..174491
/note="assembly_fragment"
174592..192060
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 49668 a 46807 c 46516 g 48562 t 507 others
ORIGIN

```

Query Match	49.6%;	Score 4465;	DB 2;	Length 192060;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 4975; Conservative	0;	Mismatches 3;	Indels 4;	Gaps 2

[illegible]

QY	721	GCACGTATCACATACAGCTGTCCAAATGTCGCTAACTCCCTAGTCCACAGATGCGCTTTACA	780
Db	40881	GCACGTATCACATACAGCTGTCCAAATGTCGCTAACTCCCTAGTCCACAGATGCGCTTTACA	4094
QY	781	CTCGCTTCTGCTTCCCATCTGCTGTTGACATTTGTCAGAACCAAGAAATTAGAAATGTCGGT	840
Db	40941	CTCGCTTCTGCTTCCCATCTGCTGTTGACATTTGTCAGAACCAAGAAATTAGAAATGTCGGT	4100
QY	841	ATTTATTTGTGTGCTGAGAGACCATCATCCAGGCGCTTTTCACATTTTCAGGACCATGTTTAC	900
Db	41001	ATTTATTTGTGTGCTGAGAGACCATCATCCAGGCGCTTTTCACATTTTCAGGACCATGTTTAC	4106
QY	901	TAACTGGGCTACTTCTCCACAGGTTTGAAACCAATTTGTTTATAATTACTTATTTTGTGT	960
Db	41061	TAACTGGGCTACTTCTCCACAGGTTTGAAACCAATTTGTTTATAATTACTTATTTTGTGT	4112
QY	961	GCATGAGGTAGGCAATGTATACGTATGTATAGAGATGCATGTAGGCGCTGACCCCTCAA	1020
Db	41121	GCATGAGGTAGGCAATGTATACGTATGTATAGAGATGCATGTAGGCGCTGACCCCTCAA	4118
QY	1021	ATCATTTGAGATPCCCAACAGTAAGTCACCGAGCGTTGTAACTGTTATATGTGGAGCTG	1080
Db	41181	ATCATTTGAGATPCCCAACAGTAAGTCACCGAGCGTTGTAACTGTTATATGTGGAGCTG	4124
QY	1081	GGAGCCAAAGCTGGGTTCTCTGCAGAGACGCCAGTGCGCTTAAACATGGGACCAAGCTCT	1140
Db	41241	GGAGCCAAAGCTGGGTTCTCTGCAGAGACGCCAGTGCGCTTAAACATGGGACCAAGCTCT	4130
QY	1141	CTAGGCGCTTAAGSTATCTTTAGTTTAAAAATATATATTTCTCAGCGCGGCTGTGGGCG	1200
Db	41301	CTAGGCGCTTAAGSTATCTTTAGTTTAAAAATATATATTTCTCAGCGCGGCTGTGGGCG	4136
QY	1201	ACAGCGCTTTATPCCCAACAGTACAGAGCGCTGAGAGTATAGGAATATATACACAGCGCCAG	1260
Db	41361	ACAGCGCTTTATPCCCAACAGTACAGAGCGCTGAGAGTATAGGAATATATACACAGCGCCAG	4142
QY	1261	CTGGGGTGTCAGAGCTTGCGCCTGTTTTTTTTGTTTTTCTTTATGTGCACTGTGCTTAA	1320
Db	41421	CTGGGGTGTCAGAGCTTGCGCCTGTTTTTTTTGTTTTTCTTTATGTGCACTGTGCTTAA	4148
QY	1321	CCTCGGTGTATGTCCGTGCAGGGGTGCAGATCCCTTGAGCGCTGGACTTAAAGCACTTG	1380
Db	41481	CCTCGGTGTATGTCCGTGCAGGGGTGCAGATCCCTTGAGCGCTGGACTTAAAGCACTTG	4154
QY	1381	TGATCACGCTCCGTTACAGATGCTGTGAAATTAAGCAACGAGTGTCCCTAGAGAGCAGCC	1440
Db	41541	TGATCACGCTCCGTTACAGATGCTGTGAAATTAAGCAACGAGTGTCCCTAGAGAGCAGCC	4160
QY	1441	AGTGCCTTTAACTTCTGAGCCACCCTTCCAAACCTGCTTTTATAGACTCTTAACTTTTG	1500
Db	41601	AGTGCCTTTAACTTCTGAGCCACCCTTCCAAACCTGCTTTTATAGACTCTTAACTTTTG	4166
QY	1501	TGTATATGTGGGAACCTGAGTGGATCTTGCACTTACCAGAGTGTGTCCGCTGTAGCATCA	1560
Db	41661	TGTATATGTGGGAACCTGAGTGGATCTTGCACTTACCAGAGTGTGTCCGCTGTAGCATCA	4172
QY	1561	CTGAGCCCGTACCCACACAGACTAGTGTATACAGTTTAAAGGCAAAACCTTAAACAATACA	1620
Db	41721	CTGAGCCCGTACCCACACAGACTAGTGTATACAGTTTAAAGGCAAAACCTTAAACAATACA	4178
QY	1621	ATAGTTGGATAGAGTTTGAATATATAGTCCCTGAGCTATTTGTTAGCGTGACCTTTCCTGTC	1680
Db	41781	ATAGTTGGATAGAGTTTGAATATATAGTCCCTGAGCTATTTGTTAGCGTGACCTTTCCTGTC	4184
QY	1681	TTTACCATGTGCTGTGAGAAAGATATAAAAATGAAGACTTGAAGTCTAAGTCTGGAAACCCACA	1740
Db	41841	TTTACCATGTGCTGTGAGAAAGATATAAAAATGAAGACTTGAAGTCTAAGTCTGGAAACCCACA	4190
QY	1741	GAGCGAGGCGGAAGCCACACTCCCTAAAGATTGTTCTGAGACTTCACATACACTTCACAT	1800
Db	41901	GAGCGAGGCGGAAGCCACACTCCCTAAAGATTGTTCTGAGACTTCACATACACTTCACAT	4196
QY	1801	AATGAGTTACAAATGATATATATATTTAGTAAATTCCTTTTAAAGGTATATTTGGGAGGA	1860

|||||  
Db 41961 AATGTTACAAATGTAATTAATTAATTTAGTAATTTCTTTTGAAGGATATATGTTGGAGGGA 42020  
OY 1861 GAGATGGCTCAGCTCCAGAGACATTGCTCTTGAGAGAGACCTGATTTCAGTTCC 1920  
Db 42021 GAGATGGCTCAGCTCCAGAGACATTGCTCTTGAGAGAGACCTGATTTCAGTTCC 42080  
OY 1921 AGGACTCATATGTTGGCTCAGAGCCATCTGTAAATCCAGTTCCAGAGGGTTCCACACCCT 1980  
Db 42081 AGGACTCATATGTTGGCTCAGAGCCATCTGTAAATCCAGTTCCAGAGGGTTCCACACCCT 42140  
OY 1981 CTCTGGGCTCCACAGAGACCATACATACATACACAGACATACATGAGGCAAAACACC 2040  
Db 42141 CTCTGGGCTCCACAGAGACCATACATACATACACAGACATACATGAGGCAAAACACC 42200  
OY 2041 CATACACACATTAATTAATTAAGAACTTAAGAGTGATGTGTTGAACATTTGCT 2100  
Db 42201 CATACACACATTAATTAATTAAGAACTTAAGAGTGATGTGTTGAACATTTGCT 42260  
OY 2101 TACACATGCTATTTGAAGACATGTACAAAGCACAACCTGAAGAGGATCTGGGCTGGAG 2160  
Db 42261 TACACATGCTATTTGAAGACATGTACAAAGCACAACCTGAAGAGGATCTGGGCTGGAG 42320  
OY 2161 AGATGGCTCAGCGGTTAAGACACTGATGCTCTTCCGAAGAGGTCCTGAGTTCAAT 2220  
Db 42321 AGATGGCTCAGCGGTTAAGACACTGATGCTCTTCCGAAGAGGTCCTGAGTTCAAT 42380  
OY 2221 CCTGACACACATGTTGGCTCACAACCATCATATGAGATCTGACACCCTCTCTGCT 2280  
Db 42381 CCTGACACACATGTTGGCTCACAACCATCATATGAGATCTGACACCCTCTCTGCT 42440  
OY 2281 GCATCTGAAGACAGCTGACAGCTACAGTACTAGATATATTAATTAATTTCTTTT 2340  
Db 42441 GCATCTGAAGACAGCTGACAGCTACAGTACTAGATATATTAATTAATTTCTTTT 42500  
OY 2341 TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGATTAAGACAGGACTCAG 2400  
Db 42501 TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGATTAAGACAGGACTCAG 42560  
OY 2401 GGTGATTAATCTATCTCTGGAGTTTTTCTTCCGCTTGCAACTGGTGGACAGAG 2460  
Db 42561 GGTGATTAATCTATCTCTGGAGTTTTTCTTCCGCTTGCAACTGGTGGACAGAG 42620  
OY 2461 CCCCTTTTCATTACACAGAGGGGCTACATATTTTGAACAACACACCTGCGAT 2520  
Db 42621 CCCCTTTTCATTACACAGAGGGGCTACATATTTTGAACAACACACCTGCGAT 42680  
OY 2521 ATGTTTACTGTCCTGCTGACTATGAGACAGCGGCGCGCG - -CACACACACAC 2578  
Db 42681 ATGTTTACTGTCCTGCTGACTATGAGACAGCGGCGCGCGCGCGCACACACACAC 42740  
OY 2579 ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGT 2638  
Db 42741 ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGT 42800  
OY 2639 CAAGAAGAGGCTCCCTCAACACAGATCTTCTCTTCCCTCTTAAGAGACACAGAT 2698  
Db 42801 CAAGAAGAGGCTCCCTCAACACAGATCTTCTCTTCCCTCTTAAGAGAGCCACGAT 42860  
OY 2699 CCAAGGTGGCAGAAAGATCTACAGGGGGCAGAGGCAAGGAGGAGGCAATGGATT 2758  
Db 42861 CCAAGGTGGCAGAAAGATCTACAGGGGGCAGAGGCAAGGAGGAGGCAATGGATT 42920  
OY 2759 CCAGAGACCTACAGCAGAGGCGCAGCAAGATCCCAAGTCCAGGGCAGGAGAGTGA 2818  
Db 42921 CCAGAGACCTACAGCAGAGGCGCAGCAAGATCCCAAGTCCAGGGCAGGAGAGTGA 42980  
OY 2819 GGCCCTGTTCCGAGAGAGGCGCAGCAAGATCCCAAGTCCAAAGGCAAGGTTTATGG 2878  
Db 42981 GGCCCTGTTCCGAGAGAGGCGCAGCAAGATCCCAAGTCCAAAGGCAAGGTTTATGG 43040  
OY 2879 CAGCTCATTAAGAGTGGAGTGTGCTCAGTCAGAAAGAGAGAGAGGAGAGGCCCTT 2938  
|||||

Db 43041 CAGCTCATTAAGAGTGGAGTGTGCTGCTCCTACAGAAAGAGAGAGAGAGGCCCTT 43100  
OY 2939 GTGCCCATCTGAGCAGAGGGTCTATGCTGATAGAGAGATCTGACAGGGGTGCCAGAGGCC 2998  
Db 43101 GTGCCCATCTGAGCAGAGGGTCTATGCTGATAGAGAGATCTGACAGGGGTGCCAGAGGCC 43160  
OY 2999 ACCTGTCTGTCCCAAGGAGACCCCAAGTGTAACTGTGGCTTGGGTCTGATTTCCAGC 3058  
Db 43161 ACCTGTCTGTCCCAAGGAGACCCCAAGTGTAACTGTGGCTTGGGTCTGATTTCCAGC 43220  
OY 3059 TACAAGACCCAGAGGATCTATCCATCCCATCCAGTGGCCCTGGCCGCCACACCC 3118  
Db 43221 TACAAGACCCAGAGGATCTATCCATCCCATCCAGTGGCCCTGGCCGCCACACCC 43280  
OY 3119 CACCCTCCGACTCCGCTGCTCCTTCTAGAGGCTGAGAGGGTGGCCAGCCCTGGTGGGGT 3178  
Db 43281 CACCCTCCGACTCCGCTGCTCCTTCTAGAGGCTGAGAGGGTGGCCAGCCCTGGTGGGGT 43340  
OY 3179 GCCTACCTGCAAGGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCTCTGAAGCTGCA 3238  
Db 43341 GCCTACCTGCAAGGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCTCTGAAGCTGCA 43400  
OY 3239 GAGCCAAAGGGGGGACACGGCAGCTCAGGCTGTGACAGGCTGTGCTGGGCTTAGGTTT 3298  
Db 43401 GAGCCAAAGGGGGGACACGGCAGCTCAGGCTGTGACAGGCTGTGCTGGGCTTAGGTTT 43460  
OY 3299 CCAAGGACCTTGGGACCTACTTCCACACCCCCCAATCCATCTCTGTTGGGCCCTATCTT 3358  
Db 43461 CCAAGGACCTTGGGACCTACTTCCACACCCCCCAATCCATCTCTGTTGGGCCCTATCTT 43520  
OY 3359 CCCCTTATATGTTGAAGAAAGTTTCCCTGGGGGGGGGGGTGTGTGAGACAAAGGTCCTT 3418  
Db 43521 CCCCTTATATGTTGAAGAAAGTTTCCCTGGGGGGGGGGGTGTGTGAGACAAAGGTCCTT 43578  
OY 3419 CGGTCTCTGACGACAGCTTCCACAACTTCTTAAGATCTCCAGAGTGTGGTCTGCT 3478  
Db 43579 CGGTCTCTGACGACAGCTTCCACAACTTCTTAAGATCTCCAGAGTGTGGTCTGCT 43638  
OY 3479 TCCACACAGGTAAGGCAATTTGGTGGGACACATGTTGACACACAGTGTGGAGGGAC 3538  
Db 43639 TCCACACAGGTAAGGCAATTTGGTGGGACACATGTTGACACACAGTGTGGAGGGAC 43698  
OY 3539 AGGTCCTGCTCTCTCTGACGCTGCTGCTCTCTGAGCAGCTTGTATTAAGTTTGG 3598  
Db 43699 AGGTCCTGCTCTCTCTGACGCTGCTGCTCTCTGAGCAGCTTGTATTAAGTTTGG 43758  
OY 3599 GGTGAGTGAAGTGTCTGAAACTCTGAAAGAGCAAGCAAGCAGAGCTGTCTTGGGC 3658  
Db 43759 GGTGAGTGAAGTGTCTGAAACTCTGAAAGAGCAAGCAAGCAGAGCTGTCTTGGGC 43818  
OY 3659 CTTCAATGAAGAGTTCACAGACCCCTTCTCTGTATGTCACCTTCTCTCATCTGTGT 3718  
Db 43819 CTTCAATGAAGAGTTCACAGACCCCTTCTCTGTATGTCACCTTCTCTCATCTGTGT 43878  
OY 3719 AGATTCCCTGGGACCAAGGTGCTCTGAGACTGAGATTCTTCAATTAATTCAGGACA 3778  
Db 43879 AGATTCCCTGGGACCAAGGTGCTCTGAGACTGAGATTCTTCAATTAATTCAGGACA 43938  
OY 3779 GTCTGAGACTTGGACCTCGTGTCTGATTTACTACTTCTCTGCTGGCTGCTCATTTCTGT 3838  
Db 43939 GTCTGAGACTTGGACCTCGTGTCTGATTTACTACTTCTCTGCTGGCTGCTCATTTCTGT 43998  
OY 3839 GTTCATGCTTACACATCTGAAGTGTCTTCTTGTGTACCAATTCCTTGACACTCTGG 3898  
Db 43999 GTTCATGCTTACACATCTGAAGTGTCTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 44058  
OY 3899 GAGTGTGATCTTGTGGCAGATGTATCTCTGGGATGTAACTGTGACGACACAGAGAGAGG 3958  
Db 44059 GAGTGTGATCTTGTGGCAGATGTATCTCTGGGATGTAACTGTGACGACACAGAGAGAGG 44118  
OY 3959 GAGAGTGTAGAGGCTGTGCTAGGCGCTTATTAAGCTGTGACATCACCCCTTCTTAGAA 4018  
Db 44119 GAGAGTGTAGAGGCTGTGCTAGGCGCTTATTAAGCTGTGACATCACCCCTTCTTAGAA 44178  
|||||

QY	4019	ATGCCCTTCATTTTTCGTTACCAATGATATTTATATGAGAGTGGGAGTGAAGC	4078
DB	44179	ATGGCCCCCTCATTTTTCGTTACCAATGATATTTATATGAGAGTGGGAGTGAAGC	44238
QY	4079	CAAACTGCGCCGAAAGTTGGGACATCTACAGACCAAGTTATCTGCTAGAAATCCCC	4138
DB	44239	CAAACTGCGCCGAAAGTTGGGACATCTACAGACCAAGTTATCTGCTAGAAATCCCC	44298
QY	4139	TCTCAGTTGAGTTGGGAGAAATCTGCTTGGGGGCTTCAGGTTCTGGTTAGCAGAGG	4198
DB	44299	TCTCAGTTGAGTTGGGAGAAATCTGCTTGGGGGCTTCAGGTTCTGGTTAGCAGAGG	44358
QY	4199	GATCCTTTGATPAGGGATGACCTAGTCTATGCTTACTACATCTCCGTCAGTTAA	4258
DB	44359	GATCCTTTGATPAGGGATGACCTAGTCTATGCTTACTACATCTCCGTCAGTTAA	44418
QY	4259	ACCTGGAAGTAAACCCAGCGACGCGCCAGATTTCTACAGTTGTACCCAGAACAA	4318
DB	44419	ACCTGGAAGTAAACCCAGCGACGCGCCAGATTTCTACAGTTGTACCCAGAACAA	44478
QY	4319	CAAGACATGATATGCAAGATGAGTGTGCTGGGGAGAAAGAACTTAACCCCCCAA	4378
DB	44479	CAAGACATGATATGCAAGATGAGTGTGCTGGGGAGAAAGAACTTAACCCCCCAA	44538
QY	4379	AGGCCACAGTTCCGTTCCCTAGTTACATGCTCAGATGAGTGTGCTTACTATGGGC	4438
DB	44539	AGGCCACAGTTCCGTTCCCTAGTTACATGCTCAGATGAGTGTGCTTACTATGGGC	44598
QY	4439	TCTGAGTTGCTGATCAAGCATGAGTGTGCTGATGCTGATATTAATCTGAGC	4498
DB	44599	TCTGAGTTGCTGATCAAGCATGAGTGTGCTGATGCTGATATTAATCTGAGC	44658
QY	4499	ACTTGGAGGCTGAAGCAGAGATGCTATATGTTTGGGCGACGCTCAGATATGAGC	4558
DB	44659	ACTTGGAGGCTGAAGCAGAGATGCTATATGTTTGGGCGACGCTCAGATATGAGC	44718
QY	4559	GAGACTTTGCTTTTAAAGAAAAATGAAAGCCAGCAGTGTGCGACAGCCTTTATCC	4618
DB	44719	GAGACTTTGCTTTTAAAGAAAAATGAAAGCCAGCAGTGTGCGACAGCCTTTATCC	44778
QY	4619	CAGCACTTGGGAGCGAAGACAGCAGATTTCTGAGTCAAGGCCAGCCTGCTATAGA	4678
DB	44779	CAGCACTTGGGAGCGAAGACAGCAGATTTCTGAGTCAAGGCCAGCCTGCTATAGA	44838
QY	4679	GTCAGTTCCAGACAGCGGCTACACAGAAACCCGTTTGAAGAAACCAAGAAAC	4738
DB	44839	GTCAGTTCCAGACAGCGGCTACACAGAAACCCGTTTGAAGAAACCAAGAAAC	44898
QY	4739	AAAAAACAACCAAAACCAAAACCAAAACCAAAACCAAAACCTCTCATCTCATCTC	4798
DB	44899	AAAAAACAACCAAAACCAAAACCAAAACCAAAACCAAAACCTCTCATCTCATCTC	44958
QY	4799	TCTAGGCTGTCTGTCTGATGCTGATGAGTTGGGACCTTCAGACTTATATATATAG	4858
DB	44959	TCTAGGCTGTCTGTCTGATGCTGATGAGTTGGGACCTTCAGACTTATATATATAG	45018
QY	4859	CCTTTTATCAGTGGTCAAGACAGAAAGTTTTCGCTGGGACACATGGGACCTTGA	4918
DB	45019	CCTTTTATCAGTGGTCAAGACAGAAAGTTTTCGCTGGGACACATGGGACCTTGA	45078
QY	4919	GAAGACTACTCTTGGCAGCCCAAAATTTCTGGGAAAGCTTCTGGAAGAGTGTGCCG	4978
DB	45079	GAAGACTACTCTTGGCAGCCCAAAATTTCTGGGAAAGCTTCTGGAAGAGTGTGCCG	45138
QY	4979	AT 4980	
DB	45139	AT 45140	

RESULT 3  
AC132885  
LOCUS AC132885 170074 bp DNA linear HMG 12-MAR-2003

DEFINITION Mus musculus clone RP24-263015, WORKING DRAFT SEQUENCE. 9 unordered pieces.  
AC132885  
VERSION AC132885.3 GI:28927760  
KEYWORDS HTGS\_PHASE1, HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
1 (bases 1 to 170074)  
AUTHORS Birren, B., Nusbaum, C. and Lander, E.  
TITLE Mus musculus, clone RP24-263015  
JOURNAL Unpublished  
2 (bases 1 to 170074)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fardo, S., Ferrel, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 170074)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corbin, B., DeRellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fardo, S., Ferrel, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachevka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Studts, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Mar 12, 2003 this sequence version replaced gi:28416187. All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L27142  
Center clone name: 263.O.15  
Summary Statistics  
Sequencing vector: n/a; 100% of reads



QY	3853	CATCTGAAATGGTTCTTTTGTCACCATTCCTCCGTACACTCTCTGGAGGTGTAATCTT	3912
Db	1139	CATCTGAAATGGTTCTTTTGTCACCATTCCTCCGTACACTCTCTGGAGGTGTAATCTT	1198
QY	3913	GGCCACATGTATCCTGGGATGTAACTCTCAGCCACACAGAGAGAGGGGAGATCAGAGC	3972
Db	1199	GGCCACATGTATCCTGGGATGTAACTCTCAGCCACACAGAGAGAGGGGAGATCAGAGC	1258
QY	3973	TGTGTCTAGGCCCTATTAGGCTCGAGACATCACCCCTTTCCTAATAATGGCCCTCCAT	4032
Db	1259	TGTGTCTAGGCCCTATTAGGCTCGAGACATCACCCCTTTCCTAATAATGGCCCTCCAT	1318
QY	4033	TTTGTGGTTACCATATCTATTTTATATCAGAGTGGCAGTGAAGCCAACTCTCCAGA	4092
Db	1319	TTTGTGGTTACCATATCTATTTTATATCAGAGTGGCAGTGAAGCCAACTCTCCAGA	1378
QY	4093	AGTTGGGACATCACTCAGACCAGAGTTATCGCTCAGAGAAATCCCCGTGCCTGAGGTT	4152
Db	1379	AGTTGGGACATCACTCAGACCAGAGTTATCGCTCAGAGAAATCCCCGTGCCTGAGGTT	1438
QY	4153	GGGGAATCTGCTCTG6GGGCTTCCAGTCTTGTTAGCAGAGAGGATTCCTTTGATA	4212
Db	1439	GGGGAATCTGCTCTG6GGGCTTCCAGTCTTGTTAGCAGAGAGGATTCCTTTGATA	1498
QY	4213	GGGCATACCTAGCTATGGTGTAGTACATTCCTGTCCAGTTAAAGCTGGACTATAA	4272
Db	1499	GGGCATACCTAGCTATGGTGTAGTACATTCCTGTCCAGTTAAAGCTGGACTATAA	1558
QY	4273	CCCAAGGAGGCCCCAGGATTCCTACAGTTGTATCCCAAGAACACAGACATAGATA	4332
Db	1559	CCCAAGGAGGCCCCAGGATTCCTACAGTTGTATCCCAAGAACACAGACATAGATA	1618
QY	4333	TGCAGAGATAGTATGCTGGGAGAGAAAGAACTTAAACCCCCCAAGGCCACAGGTT	4392
Db	1619	TGCAGAGATAGTATGCTGGGAGAGAGAACTTAAACCCCCCAAGGCCACAGGTT	1678
QY	4393	CGTTCCCTAGTTCACAAATGCCAGATATAGTCTATGCTACTGTGGCTGTGAGTTGGTAGC	4452
Db	1679	CGTTCCCTAGTTCACAAATGCCAGATATAGTCTATGCTACTGTGGCTGTGAGTTGGTAGC	1738
QY	4453	TACACGATGATGATGTTCAATGTGTAGTGTATATCTGATGAGCACTTGGAGGCTGA	4512
Db	1739	TACACGATGATGATGTTCAATGTGTAGTGTATATCTGAGCACTTGGAGGCTGA	1798
QY	4513	AGCAGGAGGATTCGTATGTTTGAAGCCAGCCGAGCTATAGAGCGAGATTTGTCTT	4572
Db	1799	AGCAGGAGGATTCGTATGTTTGAAGCCAGCCGAGCTATAGAGCGAGATTTGTCTT	1858
QY	4573	AAGAAAAAATGAAGCCACAGATGTTGGCAGACGCCCTTTAATCCAGCACTTGGGAGG	4632
Db	1859	AAGAAAAAATGAAGCCACAGATGTTGGCAGACGCCCTTTAATCCAGCACTTGGGAGG	1918
QY	4633	CAGAAAGCAGCAGATTTCTGAGTTCAGAGCCAGCCTGCTATAGAGTGAATCCAGAC	4692
Db	1919	CAGAAAGCAGCAGATTTCTGAGTTCAGAGCCAGCCTGCTATAGAGTGAATCCAGAC	1978
QY	4693	AGCCAGGCTTACACAGAGAAACCTCTGTTTGAAGAAACAGAAAAACAAACAAAA	4752
Db	1979	AGCCAGGCTTACACAGAGAAACCTCTGTTTGAAGAAACAGAAAAACAAACAAAA	2038
QY	4753	CAAAACAAAACCCAAACCCAAACCCAAACCTCTATCTCTATCTCTAGGCTGTGCT	4812
Db	2039	CAAAACAAAACCCAAACCCAAACCCAAACCTCTATCTCTATCTCTAGGCTGTGCT	2098
QY	4813	GTTAGAGTGTAGAGTTTGGGACCTTCAGATTTATATAAATAGGCTTTTATCACTG	4872
Db	2099	GTTAGAGTGTAGAGTTTGGGACCTTCAGATTTATATAAATAGGCTTTTATCACTG	2158
QY	4873	GTCAGAGACGAAAGTTTCACTCTGGGACACAGTGGAGCCGTGAGAAATCACTCTTG	4932
Db	2159	GTCAGAGACGAAAGTTTCACTCTGGGACACAGTGGAGCCGTGAGAAATCACTCTTG	2218

QY	4933	CCAGCCCCAAATTTCTGGGAAGCCTCCCGAGAGAGTGTCTCCGAT	4980
Db	2219	CCAGCCCAAAATTTCTGGGAGGCTTCCTGGAGAGTGTCTCCGAT	2266
RESULT 4			
LOCUS	BY067438/c		
DEFINITION	BY067438	779 bp	DNA
ACCESSION	5212P60465FAL.T0	CZECCHI/El	Mus musculus STS genomic, sequence tagged site.
VERSION	BY067438		
KEYWORDS	BY067438.1	GI:31183233	
SOURCE	STS		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 779)		
TITLE	Wade,C.M., Kulbokas,E.J. Iir, Kirby,A.W., Zody,M.C., Mullikin,J.C., Lander,E.S., Lindblad-Toh,K. and Daly,M.J		
JOURNAL	The mosaic structure of variation in the laboratory mouse genome		
MEDLINE	Nature 420 (6915), 574-578 (2002)		
PUBMED	22354684		
COMMENT	12466852		
	Contact: Kerstin Lindblad-Toh		
	Whitehead Institute for Biomedical Research, Center for Genome Research		
	320 Charles Street, Cambridge, MA 02141, USA		
	Tel: 6172521477		
	Fax: 6172580903		
	Email: kerslin@genome.wi.mit.edu		
	Primer A: None		
	Primer B: None		
	STS size: 779		
	Protocol:		
	WGS-discovey: Paired-end low-coverage whole genome shotgun reads were generated from 129S1/SVIMJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the MGSCV3 C57BL/6J assembly and SNP detection was carried out by SSRHA-SNP. 225,000 reads were annotated		
	as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.		
FEATURES	Location/Qualifiers		
source	1..779		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="CZECCHI/El"		
	/db_xref="taxon:10090"		
	/map="1 31-720 75128449-75127762"		
	/clone_idb="CZECCHI/El"		
	<1..>779		
STS	162 a	230 c	153 g
BASE COUNT			234 t
ORIGIN			
Query Match	2.0%;	Score 180;	DB 11;
Best Local Similarity	100.0%;	Pred. No. 8.6e-87;	Length 779;
Matches 180;	Conservative 0;	Mismatches 0;	Indels 0;
		Gaps 0;	
QY	4557	TGCTGGCACAGCCCTTTAATCCAGACACTTTGGGAGCAGACAGAGGATTTCTGAGTT	4656
Db	678	TGCTGGCACAGCCCTTTAATCCAGACACTTTGGGAGCAGACAGAGGATTTCTGAGTT	619
QY	4657	CAAGCCAGCCCTGCTCTATAGAGTGAATTCACAGACAGCCAGGAGCTACAGAGAAACCC	4716
Db	618	CAAGCCAGCCCTGCTCTATAGAGTGAATTCACAGACAGCCAGGAGCTACAGAGAAACCC	559
QY	4717	TGTTTGAATAACAGAAACAAACAAACAAACAAACAAACAAACCCAAACCCAAAC	4776
Db	558	TGTTTGAATAACAGAAACAAACAAACAAACAAACAAACAAACCCAAACCCAAAC	499



RESULT 5  
AC101018  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC101018 149109 bp DNA linear HTG 21-AUG-2002  
Rattus norvegicus clone RP32-328P7, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
9 unordered pieces.  
AC101018  
AC101018.2 GI:22381320  
HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Chordata; Sciurognathi; Muridae; Murinae;  
Rattus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 149109)  
Birren,B., Nusbaum,C. and Lander,E.  
Rattus norvegicus, clone RP32-328P7  
Unpublished  
2 (bases 1 to 149109)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,  
Brown,A., Camarata,J., Campolano,A., Chang,U., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferrira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,  
Hagos,B., Heatord,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,  
Macanar,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunhahng,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 149109)  
Birren,B., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarata,J., Chang,U., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferrira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunhahng,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thoman,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (31-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:117059792.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITB  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L14794  
Center Clone name: 328\_P-7  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 46829: contig of 46829 bp in length  
\* 46830 46929: gap of 100 bp  
\* 46930 49621: contig of 2692 bp in length  
\* 49622 49721: gap of 100 bp  
\* 49722 70009: contig of 20288 bp in length  
\* 70010 70109: gap of 100 bp  
\* 70110 77196: contig of 7087 bp in length  
\* 77197 77296: gap of 100 bp  
\* 77297 99453: contig of 22157 bp in length  
\* 99454 99553: gap of 100 bp  
\* 99554 102879: contig of 3326 bp in length  
\* 102880 102979: gap of 100 bp  
\* 102980 120352: contig of 17373 bp in length  
\* 120353 120452: gap of 100 bp  
\* 120453 131546: contig of 11094 bp in length  
\* 131547 131646: gap of 100 bp  
\* 131647 149109: contig of 17463 bp in length.  
\*  
Location/Qualifiers  
1..149109  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="RP32-328P7"  
/clone\_11b="RPC1-32 Male Rat BAC"  
BASE COUNT 39036 a 34316 c 35185 g 39719 t 853 others  
ORIGIN  
Query Match 1.1%; Score 96; DB 2; Length 149109;  
Best Local Similarity 100.0%; Pred. No. 1.2e-40;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5953 GAGTTTACAGAGGTAAAGCCATTTATATCCAGACATGCTCAAAATCAAG 6012  
Db 102361 GAGTTTACAGAGGTAAAGCCATTTATATCCAGACATGCTCAAAATCAAG 102420  
QY 6013 TGCAGAGAGATTAGCTGAGAGATGGGGCTGTCA 6048  
Db 102421 TGCAGAGAGATTAGCTGAGAGATGGGGCTGTCA 102456  
RESULT 6  
AC115173  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC115173 193472 bp DNA linear HTG 15-NOV-2002  
Rattus norvegicus clone CH230-286017, WORKING DRAFT SEQUENCE, 3  
unordered pieces.  
AC115173  
AC115173.4 GI:25006753  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 193472)  
Muzny,D., Marie., Metzker,M., Lee., Abramzon,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Ayaga,A., Ayodele,I.M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,



Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleeland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Day-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregioris,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,D., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Huliyk,S., Hume,J., Idelblid,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindarcne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawliny,S., Mcleod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Natr,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwoekelimeh,O., Okmonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Polidexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Qutroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabot,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usman,I., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Unpublished  
2 (bases 1 to 193472)  
Direct Submission  
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 193472)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23678661.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-286017  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 185211 bases at least Q40  
Consensus quality: 186794 bases at least Q30  
Consensus quality: 188105 bases at least Q20  
Estimated insert size: 184863; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 190920: contig of 190920 bp in length  
\* 190921 191020: gap of unknown length  
\* 191021 192270: contig of 1250 bp in length  
\* 192271 192370: gap of unknown length  
\* 192371 193472: contig of 1102 bp in length.  
Location/Qualifiers  
1.193472  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-286017"  
1.1118  
/note="wgs\_end\_extension  
clone\_end:t7"  
4580.5464  
/note="clone\_boundary  
clone\_end:t7"  
site:  
end\_sequence:RXAHM93TVB"  
185119.186013  
/note="clone\_boundary  
clone\_end:sp6  
site:  
end\_sequence:RXAHM93TVB"  
187253.188705  
/note="wgs\_end\_extension  
clone\_end:t7"  
189469.190920  
/note="wgs\_end\_extension  
clone\_end:t7"  
BASE COUNT 51460 a 44772 c 44401 g 48651 t 4188 others  
ORIGIN  
Query Match 1.1%; Score 96; DB 2; Length 193472;  
Best local Similarity 100.0%; Pred. No. 1.2e+40;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5953 GAGTTTCAAGAGGTAAGGCGCAAGCCATTATTCACAAAGCATCAATCAAG 6012  
|||||  
DB 125203 GAGTTTCAAGAGGTAAGGCGCAAGCCATTATTCACAAAGCATCAATCAAG 125262  
|||||  
QY 6013 TGCAGAGAGATTTAGCTGAGAGATGGGGCTGCA 6048  
|||||  
DB 125263 TGCAGAGAGATTTAGCTGAGAGATGGGGCTGCA 125298  
|||||

RESULT 7  
AC136730/c 145489 bp DNA linear HTG 22-FEB-2003  
LOCUS AC136730.3  
DEFINITION Mus musculus clone RP23-76E22, WORKING DRAFT SEQUENCE, 4 unordered pieces  
ACCESSION AC136730 GI:28467347  
VERSION AC136730.3  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 145489)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP23-76E22  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145489)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barne,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fardo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,T., Mathews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenwa,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 145489)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barne,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Deatellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fardo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,T., Manning,J., Mathews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenwa,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Roman,J., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Feb 22, 2003 this sequence version replaced gi:28412006. All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WtBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information

Center project name: L26307  
Center clone name: 76\_E-22  
----- Summary Statistics  
Sequencing vector: Plasmid: n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 144239 bases at least Q40  
Consensus quality: 144835 bases at least Q30  
Consensus quality: 145007 bases at least Q20  
Insert size: 145000; agarose-fp  
Insert size: 145189; sum-of-ctnigs  
Quality coverage: 8.6 in Q20 bases; sum-of-ctnigs  
Quality coverage: 8.6 in Q20 bases; sum-of-ctnigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 14246: contig of 14246 bp in length  
\* 14247 14346: gap of 100 bp  
\* 14347 16056: contig of 1710 bp in length  
\* 16057 16156: gap of 100 bp  
\* 16157 111985: contig of 95829 bp in length  
\* 111986 112085: gap of 100 bp  
\* 112086 145489: contig of 33404 bp in length.  
\* location/Qualifiers  
1. 145489  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-76E22"  
/clone.lib="RCI-23 Female Mouse BAC"  
1. 14246  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
14347. 16056  
/note="assembly\_fragment"  
16157. 111985  
/note="assembly\_fragment"  
112086. 145489  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"  
BASE COUNT 40514 a 34404 c 33449 g 36801 t 321 others  
ORIGIN  
Query Match 0.98; Score 82; DB 2; Length 145489;  
Best Local Similarity 100.0%; Pred. No. 5.9e-33;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4638 GCAGGAGATTTCGTGATTCAGGCGCTGCTTACAGTGTCCAGACGCA 4697  
Db 87852 GCAGGAGATTTCGTGATTCAGGCGCTGCTTACAGTGTCCAGACGCA 87793  
QY 4698 GGGCTACACAGAGAAACCTGT 4719  
Db 87792 GGGCTACACAGAGAAACCTGT 87771  
-----  
RESULT 8  
AC115850 154704 bp DNA linear HTG 06-MAR-2003  
LOCUS AC115850.4  
DEFINITION Mus musculus clone RP24-252010, WORKING DRAFT SEQUENCE, 4 unordered pieces  
ACCESSION AC115850  
VERSION AC115850.4 GI:28867059  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS  
1 (bases 1 to 154704)  
Birren,B., Nusbaum,C. and Lander,E.  
MUS musculus, clone RP24-252010  
JOURNAL  
TITLE  
2 (bases 1 to 154704)  
Anderson,S., Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Bouckigalter,B., Brown,A., Camarata,J., Bloom,T., Boguslavsky,L.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,  
Landers,T., Lenocksky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Mathews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,  
Mihova,C., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,M., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strausman,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
JOURNAL  
TITLE  
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE  
AUTHORS  
3 (bases 1 to 154704)  
Birren,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Bouckigalter,B., Camarata,J., Chang,J.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Halez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
JOURNAL  
TITLE  
Submitted (06-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT  
On Mar 6, 2003 this sequence version replaced gi:28631266.  
All repeats were identified using RepeatMasker:  
Smith,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WtBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center Project name: L24674  
Center Clone name: 252.O.10  
----- Summary Statistics  
Sequencing vector: Plasmid, n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 153993 bases at least Q40  
Consensus quality: 154237 bases at least Q30  
Consensus quality: 154353 bases at least Q20  
Insert size: 157000; agarose-ff  
Insert size: 154404; sum-of-contigs  
Quality coverage: 7.4 in Q20 bases; agarose-ff  
Quality coverage: 7.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1. 2016: contig of 2016 bp in length  
\* 2017 2116: gap of 100 bp  
\* 2117 44832: contig of 42716 bp in length  
\* 44833 44932: gap of 100 bp  
\* 44933 113063: contig of 68131 bp in length  
\* 113064 113163: gap of 100 bp  
\* 113164 154704: contig of 41541 bp in length.  
\* Location/Qualifiers  
1. 154704  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP24-252010"  
/clone\_1ib="RPCI-24 Male Mouse BAC"  
1. 2016  
/note="assembly\_fragment"  
2117. 44832  
/note="assembly\_fragment"  
44933. 113063  
/note="assembly\_fragment"  
113164. 154704  
/note="assembly\_fragment"  
clone\_end:f7  
vector\_side:right"  
BASE COUNT 39246 a 35851 c 36618 g 42687 t 302 others  
ORIGIN  
Query Match 0.9%: Score 82; DB 2; Length 154704;  
Best Local Similarity 100.0%: Pred. No. 5 9e-33;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4638 GCAGGCAGATTCTGAGTTCAAGCCAGCCTGCTATAGAGTGAGTTCAGACACCA 4697  
Db 150968 GCAGGCAGATTCTGAGTTCAAGCCAGCCTGCTATAGAGTGAGTTCAGACACCA 151027  
QY 4698 GGGCTACACAGAGAACCCCTGT 4719  
Db 151028 GGGCTACACAGAGAACCCCTGT 151049

RESULT 9  
AC120393/c 173892 bp DNA linear HTG 23-APR-2003  
LOCUS  
DEFINITION  
MUS musculus clone RP24-312B12, WORKING DRAFT SEQUENCE, 21  
unordered pieces.  
ACCESSION  
AC120393  
VERSION  
AC120393.3 GI:30017810  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
SOURCE  
MUS musculus (house mouse)  
ORGANISM  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS  
1 (bases 1 to 173892)  
Birren,B., Nusbaum,C. and Lander,E.  
MUS musculus, clone RP24-312B12  
JOURNAL  
TITLE  
2 (bases 1 to 173892)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campolongo, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gorda, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarty, M., McKean, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (06-May-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
(bases 1 to 173892)  
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Mclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Menes, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rogov, P., Rachupka, A., Ramasamy, V., Raymond, C., Retta, R., Rise, C., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smlth, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-Apr-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 17, 2003 this sequence version replaced g1:28650112.  
All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web Site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L25953  
Center clone name: 312\_B\_12  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 236: contig of 236 bp in length  
\* 237 336: gap of 100 bp  
\* 337 703: contig of 367 bp in length

## FEATURES

## source

\* 704 803: gap of 100 bp  
\* 804 3285: contig of 2482 bp in length  
\* 3286 3385: gap of 100 bp  
\* 3386 5646: contig of 2261 bp in length  
\* 5647 5747: gap of 100 bp  
\* 5747 8535: contig of 2789 bp in length  
\* 8536 8636: gap of 100 bp  
\* 8636 11961: contig of 3326 bp in length  
\* 11962 12061: gap of 100 bp  
\* 12062 14947: contig of 2886 bp in length  
\* 14948 15048: gap of 100 bp  
\* 15048 17510: contig of 2463 bp in length  
\* 17511 17610: gap of 100 bp  
\* 17611 20267: contig of 2657 bp in length  
\* 20268 20367: gap of 100 bp  
\* 20368 25647: contig of 5280 bp in length  
\* 25648 25747: gap of 100 bp  
\* 25748 30334: contig of 4587 bp in length  
\* 30335 30434: gap of 100 bp  
\* 30435 38139: contig of 7705 bp in length  
\* 38140 38240: gap of 100 bp  
\* 38240 50444: contig of 12205 bp in length  
\* 50445 50544: gap of 100 bp  
\* 50545 60889: contig of 10345 bp in length  
\* 60890 60990: gap of 100 bp  
\* 60990 71937: contig of 10947 bp in length  
\* 71937 72037: gap of 100 bp  
\* 72037 88642: contig of 16606 bp in length  
\* 88643 88742: gap of 100 bp  
\* 88743 101771: contig of 10209 bp in length  
\* 101772 101872: gap of 100 bp  
\* 101872 115509: contig of 13638 bp in length  
\* 115510 115609: gap of 100 bp  
\* 115610 138341: contig of 22732 bp in length  
\* 138342 138441: gap of 100 bp  
\* 138442 159683: contig of 21242 bp in length  
\* 159684 159783: gap of 100 bp  
\* 159784 173892: contig of 14109 bp in length.  
Location/Qualifiers  
1..173892  
/organism="Mus musculus"  
/db\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP24-312B12"  
/clone\_11b="RPCT-24 Male Mouse BAC"  
1..236  
/note="assembly\_fragment"  
337..703  
/note="assembly\_fragment"  
804..3285  
/note="assembly\_fragment"  
3386..5646  
/note="assembly\_fragment"  
5747..8535  
/note="assembly\_fragment"  
8636..11961  
/note="assembly\_fragment"  
12062..14947  
/note="assembly\_fragment"  
15048..17510  
/note="assembly\_fragment"  
17611..20267  
/note="assembly\_fragment"  
20368..25647  
/note="assembly\_fragment"  
25748..30334  
/note="assembly\_fragment"  
30435..38139  
/note="assembly\_fragment"  
38240..50444  
/note="assembly\_fragment"  
50545..60889  
/note="assembly\_fragment"

```

misc_feature      60990..71936
                  /note="assembly-fragment"
misc_feature      72037..88642
                  /note="assembly-fragment"
misc_feature      88743..101771
                  /note="assembly-fragment"
misc_feature      101872..115509
                  /note="assembly-fragment"
misc_feature      115610..138341
                  /note="assembly-fragment"
misc_feature      138442..159683
                  /note="assembly-fragment"
misc_feature      159784..173892
                  /note="assembly-fragment"
                  clone_end:T7
                  vector_side:right"
BASE COUNT      41036 a 45314 c 42508 g 43019 t 2015 others
ORIGIN
Query Match      0.9%; Score 82; DB 2; Length 173892;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4593 GCAGTGTGGCACACGCTTTAATCCACGACTTGGAGGCAGACGACGAGATTTCGTG 4652
         |||||||
Db      146185 GCAGTGTGGCACACGCTTTAATCCACGACTTGGAGGCAGACGACGAGATTTCGTG 146126
         |||||||

QY      4653 AGTTCAAGCCGCTGTCTTA 4674
         |||||||
Db      146125 AGTTCAAGCCGCTGTCTTA 146104
         |||||||

RESULT 10
AC139041/c      189241 bp      DNA      linear      HTG 17-MAR-2003
LOCUS
DEFINITION
SEQUENCE: 6 unordered pieces.
AC139041
AC139041.3 GI:28976013
HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE
MUS musculus (house mouse)
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 189241)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 14, clone RP23-283B12
Unpublished
2 (bases 1 to 189241)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lander,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Menes,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachpka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 189241)

```

```

AUTHORS
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lander,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Menes,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachpka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 17, 2003 this sequence version replaced gi:28603968.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: 283_B-12
Center clone name: 283_B-12

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187878 bases at least Q40
Consensus quality: 188076 bases at least Q40
Consensus quality: 188202 bases at least Q20
Insert size: 175000; agarose-fp
Quality coverage: 13.1 in Q20 bases; agarose-fp
Quality coverage: 12.1 in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 48250: contig of 48250 bp in length
* 48251 48350: gap of 100 bp
* 48351 53618: contig of 5268 bp in length
* 53619 53718: gap of 100 bp
* 53719 62174: contig of 8456 bp in length
* 62175 62274: gap of 100 bp
* 62275 109247: contig of 46973 bp in length
* 109248 109347: gap of 100 bp
* 109348 148402: contig of 39055 bp in length
* 148403 148502: gap of 100 bp
* 148503 189241: contig of 40739 bp in length.
Location/Qualifiers
1..189241
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/map="14"

```

## FEATURES

source

```

misc_feature      /clone="Rp23-283812"
                  .48250
                  /clone_id="RPCT-23 Female Mouse BAC"
misc_feature      /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:left"
misc_feature      48351..53618
                  /note="assembly_fragment"
                  53719..62174
misc_feature      /note="assembly_fragment"
                  62275..109247
misc_feature      /note="assembly_fragment"
                  109348..148402
misc_feature      /note="assembly_fragment"
                  148503..189241
misc_feature      /note="assembly_fragment"
                  clone_end:T7
                  vector_side:right"
BASE COUNT      48741 a 45049 c 45485 g 49461 t 505 others
ORIGIN
Query Match      0.9%; Score 82; DB 2; Length 189241;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4638 GCAGGACATTTCGAGTTCAGGCGACCTGCTCTATAGAGTCCAGGACGCCA 4697
Db 143086 GCAGGACATTTCGAGTTCAGGCGACCTGCTCTATAGAGTCCAGGACGCCA 143027
OY 4698 GGCGTACACAGAGAACCCTGT 4719
Db 143026 GGCGTACACAGAGAACCCTGT 143005

RESULT 11
AC131335/c 190776 bp DNA linear HTG 08-MAR-2003
LOCUS Mus musculus clone Rp23-32D9, WORKING DRAFT SEQUENCE, 9 unordered
DEFINITION pieces.
ACCESSION AC131335
VERSION AC131335.4 GI:28882340
KEYWORDS HTG; PHASE1; HTGS; DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 190776)
Biren,B., Nusbäum,C. and Lander,E.
Mus musculus, clone Rp23-32D9
Unpublished
2 (bases 1 to 190776)
Biren,B., Nusbäum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,J., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Plerre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
TITLE
JOURNAL
REFERENCE
3 (bases 1 to 190776)

```

```

AUTHORS
Biren,B., Nusbäum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafer,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Plerre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tefaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (08-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 8, 2003 this sequence version replaced gi:25956385.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24187
Center clone name: 32.D_9
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 188840 bases at least 40
Consensus quality: 189359 bases at least 30
Consensus quality: 189634 bases at least 20
Insert size: 18200; agarose-fp
Insert size: 189976; sum-of-contigs
Quality coverage: 8.4 in Q20 bases; agarose-fp
Quality coverage: 8.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 34490: contig of 34490 bp in length
* 34491 34590: gap of 100 bp
* 34591 38121: contig of 3531 bp in length
* 38122 38222: gap of 100 bp
* 38222 43662: contig of 5441 bp in length
* 43663 43762: gap of 100 bp
* 43763 45721: contig of 5809 bp in length
* 45722 49671: gap of 100 bp
* 49672 58834: contig of 9163 bp in length
* 58835 58934: gap of 100 bp
* 58935 77918: contig of 18984 bp in length
* 77919 78018: gap of 100 bp
* 78019 127564: contig of 43546 bp in length
* 127565 127664: gap of 100 bp
* 127665 167916: contig of 40252 bp in length
* 167917 168016: gap of 100 bp
* 168017 190776: contig of 22760 bp in length.
Location/Qualifiers

```

```
source
1. .190776
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-32D9"
/clone_11b="RP23-32D9"
1. .34490
/feature="assembly-fragment"
clone_end:sp6
vector_side:left"
misc_feature
34591..38121
/feature="assembly-fragment"
misc_feature
38222..43662
/feature="assembly-fragment"
misc_feature
43763..49571
/feature="assembly-fragment"
misc_feature
49672..58834
/feature="assembly-fragment"
misc_feature
58935..77918
/feature="assembly-fragment"
misc_feature
78019..127564
/feature="assembly-fragment"
misc_feature
127665..167916
/feature="assembly-fragment"
misc_feature
168017..190776
/feature="assembly-fragment"
clone_end:t7
vector_side:right"
BASE COUNT 57965 a 34254 c 34054 g 63703 t 800 others
ORIGIN
Query Match 0.9%; Score 82; DB 2; Length 190776;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4638 GCAGCAGATTCTGATTCAGGCCAGCCTGCTATAGAGTTCAGAGCAGCA 4697
|||||
DB 60883 GCAGCAGATTCTGATTCAGGCCAGCCTGCTATAGAGTTCAGAGCAGCA 60824
QY 4698 GGGCTACAGAGAAACCTGT 4719
|||||
DB 60823 GGGCTACAGAGAAACCTGT 60802
RESULT 12
AC122900 203855 bp DNA linear ROD 21-SEP-2002
DEFINITION Mus musculus chromosome 15 clone RP23-39N24, complete sequence.
ACCESSION AC122900
VERSION AC122900.3 GI:23266394
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (21-SEP-2002) Genome Sequencing Center, 4444 Forest Park
```

```
COMMENT
Parkway, St. Louis, MO 63108, USA
On Sep 21, 2002 this sequence version replaced g1:22476032.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0039N24
-----
FEATURES
source
1. .203855
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP23-39N24"
BASE COUNT 61572 a 38691 c 38009 g 65583 t
ORIGIN
Query Match 0.9%; Score 82; DB 10; Length 203855;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4593 GCAGTGTGGACACGCCCTTATCCAGCACTTGGAGGACAGCAGCAGATTTCG 4652
|||||
DB 170522 GCAGTGTGGACACGCCCTTATCCAGCACTTGGAGGACAGCAGCAGATTTCG 170581
QY 4653 AGTTCAAGGCCAGCCTGTGCTTA 4674
|||||
DB 170582 AGTTCAAGGCCAGCCTGTGCTTA 170603
RESULT 13
AC127697 222185 bp DNA linear HTG 22-JUL-2002
LOCUS AC127697
DEFINITION Mus musculus chromosome UNK clone RP24-83110, WORKING DRAFT
SEQUENCE. 13 unordered pieces.
AC127697.2 GI:21913472
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 222185)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 222185)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (18-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 222185)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (22-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 21, 2002 this sequence version replaced g1:21903681.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0083110
-----
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
```

Chemistry: Dye-primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 217367 bases at least Q40  
 Consensus quality: 218557 bases at least Q30  
 Consensus quality: 219259 bases at least Q20

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      1042: contig of 1042 bp in length
*      1043      1142: gap of unknown length
*      1143      3141: contig of 1999 bp in length
*      3142      3241: gap of unknown length
*      3242      7211: contig of 3970 bp in length
*      7212      7311: gap of unknown length
*      7312      10408: contig of 3097 bp in length
*      10409      10508: gap of unknown length
*      10509      17015: contig of 6507 bp in length
*      17016      17115: gap of unknown length
*      17116      27558: contig of 10443 bp in length
*      27559      27658: gap of unknown length
*      27659      41472: contig of 13814 bp in length
*      41473      41572: gap of unknown length
*      41573      58617: contig of 17045 bp in length
*      58618      58718: gap of unknown length
*      58718      83267: contig of 24549 bp in length
*      83267      83367: gap of unknown length
*      83367      130000: contig of 46634 bp in length
*      130001      172736: contig of 42636 bp in length
*      172737      172836: gap of unknown length
*      172837      221107: contig of 48271 bp in length
*      221108      221207: gap of unknown length
*      221208      222185: contig of 978 bp in length.
  
```

FEATURES

```

source      1..222185
             /organism="Mus musculus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10090"
             /chromosome="UNK"
             /clone="RP24-83110"
misc_feature 1..1042
             /note="assembly_name:Contig25"
misc_feature 1143..3141
             /note="assembly_name:Contig26"
misc_feature 3242..7211
             /note="assembly_name:Contig27"
misc_feature 7312..10408
             /note="assembly_name:Contig28"
misc_feature 10509..17015
             /note="assembly_name:Contig29"
misc_feature 17116..27558
             /note="assembly_name:Contig30"
misc_feature 27659..41472
             /note="assembly_name:Contig31"
misc_feature 41573..58617
             /note="assembly_name:Contig32"
misc_feature 58718..83266
             /note="assembly_name:Contig33"
misc_feature 83367..130000
             /note="assembly_name:Contig34"
misc_feature 130101..172736
             /note="assembly_name:Contig35"
misc_feature 172837..221107
             /note="assembly_name:Contig36"
misc_feature 221208..222185
             /note="assembly_name:Contig37"
  
```

BASE COUNT 67728 a 41550 c 41568 g 70139 t 1200 others  
 ORIGIN

Query Match 0.9%; Score 82; DB 2; Length 222185;  
 Best Local Similarity 100.0%; Pred. No. 5,9e-33;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY      4593 GCAGTGTGCACACAGCCCTTTAAATCCAGCAGCTTGGAGGACGACGAGCATTTCTG 4652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB      26396 GCATTCGGGCGACACAGCCTTTAAATCCAGCAGCTTGGAGGACGACGAGCATTTCTG 26455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OY      4653 AGTTCAGGCCAGCAGCTGTCTA 4674
          ||||||||||||||||||||||||
DB      26456 AGTTCAGGCCAGCAGCTGTCTA 26477
  
```

```

RESULT 14
AC079490/c 226909 bp DNA linear HTG 02-SEP-2000
LOCUS      Mus musculus clone RP23-190A2, WORKING DRAFT SEQUENCE, 53 unordered
DEFINITION pieces.
ACCESSION AC079490
VERSION   AC079490.1 GI:9964855
KEYWORDS  HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 226909)
           DOE Joint Genome Institute.
           Sequencing of Mouse
           Unpublished
           2 (bases 1 to 226909)
           DOE Joint Genome Institute.
REFERENCE 2
           Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS   -----Genome Center
           Center: Joint Genome Institute
           Center Code: JGI
COMMENT    Web site: http://www.jgi.doe.gov
  
```

```

-----
Project Information
Center Project Name: 1809834
Center clone name: RP23-190A2
-----
Summary Statistics
Consensus quality: 180161 bases at least Q40
Consensus quality: 198601 bases at least Q30
Consensus quality: 203733 bases at least Q20
Estimated insert size: 209000; agarose-1p estimation
Estimated insert size: 221709; sum-of-contigs estimation
Quality coverage: 5.84 in Q20 bases; sum-of-contigs estimation.
Quality coverage: 5.84 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1120: contig of 1120 bp in length
*      1121      1220: gap of unknown length
*      1221      2574: contig of 1354 bp in length
*      2575      2674: gap of unknown length
*      2675      4062: contig of 1388 bp in length
*      4063      4162: gap of unknown length
*      4163      5666: contig of 1504 bp in length
*      5667      5767: gap of unknown length
*      5767      6895: contig of 1129 bp in length
*      6896      6995: gap of unknown length
*      6996      8025: contig of 1030 bp in length
  
```



8026 8125: gap of unknown length  
 8126 9388: contig of 1263 bp in length  
 9389 9488: gap of unknown length  
 9489 10710: contig of 1222 bp in length  
 10711 10810: gap of unknown length  
 10811 11811: contig of 1001 bp in length  
 11812 11911: gap of unknown length  
 11912 12965: contig of 1054 bp in length  
 12966 13065: gap of unknown length  
 13066 14268: contig of 1203 bp in length  
 14269 14368: gap of unknown length  
 14369 15846: contig of 1478 bp in length  
 15847 15946: gap of unknown length  
 15947 17219: contig of 1273 bp in length  
 17220 17319: gap of unknown length  
 17320 18672: contig of 1353 bp in length  
 18673 18772: gap of unknown length  
 18773 19999: contig of 1227 bp in length  
 20000 20099: gap of unknown length  
 20100 21459: contig of 1360 bp in length  
 21460 21559: gap of unknown length  
 21560 22622: contig of 1063 bp in length  
 22623 22722: gap of unknown length  
 22723 23726: contig of 1004 bp in length  
 23727 23826: gap of unknown length  
 23827 25149: contig of 1323 bp in length  
 25150 25249: gap of unknown length  
 25250 26839: contig of 1590 bp in length  
 26840 26939: gap of unknown length  
 26940 28630: contig of 1691 bp in length  
 28631 28730: gap of unknown length  
 28731 30265: contig of 1535 bp in length  
 30266 31703: contig of 1338 bp in length  
 31704 31803: gap of unknown length  
 31804 32951: contig of 1148 bp in length  
 32952 33051: gap of unknown length  
 33052 33329: contig of 2278 bp in length  
 33330 35429: gap of unknown length  
 35430 38496: contig of 3067 bp in length  
 38497 38596: gap of unknown length  
 38597 40521: contig of 1925 bp in length  
 40522 40621: gap of unknown length  
 40623 42212: contig of 1591 bp in length  
 42213 44441: contig of 2129 bp in length  
 44442 44541: gap of unknown length  
 44542 46397: contig of 1856 bp in length  
 46398 46497: gap of unknown length  
 46498 48685: contig of 2188 bp in length  
 48686 48785: gap of unknown length  
 50811 50911: contig of 2026 bp in length  
 50912 52640: gap of unknown length  
 52641 52740: gap of unknown length  
 52741 53952: contig of 1212 bp in length  
 53953 54052: gap of unknown length  
 54053 58018: contig of 3966 bp in length  
 58019 58118: gap of unknown length  
 58119 61436: contig of 3318 bp in length  
 61437 61536: gap of unknown length  
 61537 64032: contig of 2496 bp in length  
 64033 64132: gap of unknown length  
 64133 66752: contig of 2620 bp in length  
 66753 66852: gap of unknown length  
 66853 70097: contig of 3245 bp in length  
 70098 70197: gap of unknown length  
 70198 73503: contig of 3306 bp in length  
 73504 80346: contig of 6743 bp in length  
 80347 80446: gap of unknown length  
 80447 86466: contig of 6020 bp in length  
 86467 86566: gap of unknown length

86567 92278: contig of 5712 bp in length  
 92279 92378: gap of unknown length  
 92379 98374: contig of 6496 bp in length  
 98375 98875: gap of unknown length  
 98875 106050: contig of 7076 bp in length  
 106051 106150: gap of unknown length  
 106150 113625: contig of 7475 bp in length  
 113626 113725: gap of unknown length  
 113726 129075: contig of 15350 bp in length  
 129076 129175: gap of unknown length  
 129176 141800: contig of 12625 bp in length  
 141801 141900: gap of unknown length  
 141901 155581: contig of 13681 bp in length  
 155582 170325: gap of unknown length  
 170325 170424: gap of unknown length  
 170425 184681: contig of 14257 bp in length  
 184682 184781: gap of unknown length  
 184782 206159: contig of 21378 bp in length  
 206160 226909: gap of unknown length  
 226909 226909: contig of 20650 bp in length.

FEATURES  
 source  
 1..226909  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-190A2"  
 /clone\_1b="RPII mouse BAC library 23"

BASE COUNT 52802 a 57780 c 58798 g 52284 t 5245 others

ORIGIN

Query Match  
 Best local Similarity 100.0%; Pred. No. 5,9e+33;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4593 GCAGGTGGTGACACGCGCTTTAATCCAGCACTTGGAGGACAGACGAGCAGATTCTG 4652  
 DB 35815 GCAGGTGGTGACACGCGCTTTAATCCAGCACTTGGAGGACAGACGAGCAGATTCTG 35756

QY 4653 AGTTCAAGCGCAGCTGGTGCTA 4674  
 DB 35755 AGTTCAAGCGCAGCTGGTGCTA 35734

RESULT 15  
 AC141471/c  
 LOCUS AC141471 232568 bp DNA linear HTG 18-MAR-2003  
 DEFINITION Mus musculus chromosome UNK clone RP24-326K11, WORKING DRAFT  
 AC141471  
 AC141471.2 GI:29029419  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 232568)  
 McPherson,J.D. and Waterston,R.H.  
 McPherson,J.D. and Waterston,R.H.  
 The sequence of Mus musculus clone  
 unpublished  
 2 (bases 1 to 232568)  
 McPherson,J.D. and Waterston,R.H.  
 McPherson,J.D. and Waterston,R.H.  
 Direct Submission  
 Submitted (16-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 232568)  
 McPherson,J.D. and Waterston,R.H.  
 McPherson,J.D. and Waterston,R.H.  
 Direct Submission  
 Submitted (18-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 On Mar 18, 2003 this sequence version replaced gi:28975171.  
 ----- Genome Center -----

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
-----  
Project Information  
Center project name: M\_BB0326K11  
-----  
Summary Statistics  
-----  
Sequencing vector: M13, %  
Sequencing vector: Plasmid, %  
Chemistry: Dye-terminator ET, % of reads  
Chemistry: Dye-terminator Big Dye, % of reads  
Assembly program: Phrap, version 0.990319  
Consensus quality: 220808 bases at least Q40  
Consensus quality: 22401 bases at least Q30  
Consensus quality: 223551 bases at least Q20  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----  
1 1364: contig of 1364 bp in length  
\* 1365 1464: gap of unknown length  
\* 1465 3212: contig of 1748 bp in length  
\* 3213 3312: gap of unknown length  
\* 3313 5171: contig of 1859 bp in length  
\* 5172 5272: gap of unknown length  
\* 5272 7005: contig of 1734 bp in length  
\* 7006 7106 9867: gap of unknown length  
\* 9867 9967: contig of 2762 bp in length  
\* 9968 13925: gap of unknown length  
\* 13926 14025: contig of 3958 bp in length  
\* 14026 17200: contig of 3175 bp in length  
\* 17201 17300: gap of unknown length  
\* 17301 24239: contig of 6939 bp in length  
\* 24240 24339: gap of unknown length  
\* 24340 31078: contig of 6739 bp in length  
\* 31079 31179: gap of unknown length  
\* 31179 38186: contig of 7008 bp in length  
\* 38187 38286: gap of unknown length  
\* 38287 47404: contig of 9118 bp in length  
\* 47405 47504: gap of unknown length  
\* 47505 59212: contig of 11708 bp in length  
\* 59213 59312: gap of unknown length  
\* 59313 71856: contig of 12544 bp in length  
\* 71857 71956: gap of unknown length  
\* 71957 89788: contig of 17832 bp in length  
\* 89789 89888: gap of unknown length  
\* 89889 107217: contig of 17329 bp in length  
\* 107218 107317: gap of unknown length  
\* 107318 121884: contig of 14567 bp in length  
\* 121885 121984: gap of unknown length  
\* 121985 140588: contig of 18604 bp in length  
\* 140589 140688: gap of unknown length  
\* 140689 163356: contig of 22668 bp in length  
\* 163357 163456: gap of unknown length  
\* 163457 192246: contig of 28790 bp in length  
\* 192247 192346: gap of unknown length  
\* 192347 232568: contig of 40222 bp in length.  
-----  
Location/Qualifiers  
1. .232568  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP24-326K11"  
1. .1364  
/note="assembly\_name:Contig17"

misc\_feature 1465. .3212  
/note="assembly\_name:Contig19"  
misc\_feature 3313. .5171  
/note="assembly\_name:Contig21"  
misc\_feature 5272. .7005  
/note="assembly\_name:Contig22"  
misc\_feature 7106. .9867  
/note="assembly\_name:Contig25"  
misc\_feature 9968. .13925  
/note="assembly\_name:Contig26"  
misc\_feature 14026. .17200  
/note="assembly\_name:Contig27"  
misc\_feature 17301. .24239  
/note="assembly\_name:Contig28"  
misc\_feature 24340. .31078  
/note="assembly\_name:Contig29"  
misc\_feature 31179. .38186  
/note="assembly\_name:Contig30"  
misc\_feature 38287. .47404  
/note="assembly\_name:Contig31"  
misc\_feature 47505. .59212  
/note="assembly\_name:Contig32"  
misc\_feature 59313. .71856  
/note="assembly\_name:Contig33"  
misc\_feature 71957. .89788  
/note="assembly\_name:Contig34"  
misc\_feature 89889. .107217  
/note="assembly\_name:Contig35"  
misc\_feature 107318. .121884  
/note="assembly\_name:Contig36"  
misc\_feature 121985. .140588  
/note="assembly\_name:Contig37"  
misc\_feature 140689. .163356  
/note="assembly\_name:Contig38"  
misc\_feature 163457. .192246  
/note="assembly\_name:Contig39"  
misc\_feature 192347. .232568  
/note="assembly\_name:Contig40"  
BASE COUNT 54848 a 58265 c 57016 g 60272 t 2167 others  
ORIGIN  
Query Match 0.9% Score 82: DB 2: Length 232568;  
Best Local Similarity 100.0% Pred. No. 5.9e-33;  
Matches 82: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
QY 4638 GCAGCAGATTCTGAGTTCAGGCGCTGTCTATAGAGTCAAGTCCAGACACCA 4697  
|||||  
Db 118996 GCAGCAGATTCTGAGTTCAGGCGCTGTCTATAGAGTCAAGTCCAGACACCA 118937  
QY 4698 GGGCTACACAGAGAACCCTGT 4719  
|||||  
Db 118936 GGGCTACACAGAGAACCCTGT 118915

Search completed: September 13, 2003, 21:35:34  
Job time : 31324 secs